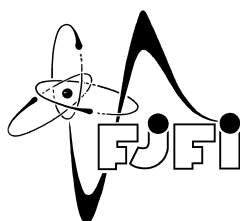


CZECH TECHNICAL UNIVERSITY IN PRAGUE  
FACULTY OF NUCLEAR SCIENCES AND PHYSICAL  
ENGINEERING



## DOCTORAL THESIS

SELECTED APPLICATION AREAS OF IMAGE  
PROCESSING: IMAGE FORENSICS AND MEDICAL  
IMAGING



PRAGUE 2017

ADAM NOVOZÁMSKÝ



## DECLARATION

---

I declare that I carried out this doctoral Thesis independently and only with the cited sources, literature and other professional sources.

*Benátky nad Jizerou, September 2017*

---

Adam Novozámský



## BIBLIOGRAFICKÝ ZÁZNAM

---

Autor	Ing. Adam Novozámský České vysoké učení technické v Praze Fakulta jaderná a fyzikálně inženýrská Katedra matematiky
Název práce	Vybrané aplikační oblasti zpracování obrazu: forenzní analýza obrazu a lékařské zobrazovací metody
Studijní program	Softwarové inženýrství
Studijní obor	Matematické inženýrství
Školitel	Ing. Stanislav Saic, CSc. Ústav teorie informace a automatizace Akademie věd České republiky, v.v.i.
Akademický rok	2017
Počet stran	153
Klíčová slova	digitální zpracování obrazu, forenzní analýza obrazu, lékařské zobrazovací metody, detekce padělků, kapslová endoskopie, videokymografie



## BIBLIOGRAPHIC ENTRY

---

Author	Ing. Adam Novozámský Czech Technical University in Prague Faculty of Nuclear Sciences and Physical Engineering Department of Mathematics
Title of Dissertation	Selected Application Areas of Image Processing: Image Forensics and Medical Imaging
Degree Programme	Software Engineering
Field of Study	Mathematical Engineering
Supervisor	Ing. Stanislav Saic, CSc. Institute of Information Theory and Automation Czech Academy of Sciences
Academic Year	2017
Number of Pages	153
Keywords	Image Processing, Image Forensics, Medical Imaging, forgery detection, capsule endoscopy, videokymography



*Science brings men nearer to God.*

— Louis Pasteur

To my son *Vojtěch*, without whom this Thesis would have been completed two years earlier.

To my lovely wife *Eva*, without whom this would have been completed three years later.

Thank God that *Alžběta* was born after the completion of the main part.

I love you all.



## ABSTRAKT

---

Tato disertační práce se zabývá dvěma oblastmi, které využívají pokročilé metody zpracování obrazové informace. Forenzní analýzou obrazu a analýzou lékařských obrazových dat. Práce nejprve uvádí do problematiky, a poté popisuje šest přiložených článků společně s autorovým přínosem. První článek představuje několik metod forenzní analýzy obrazu a videa. Ty byly navrženy tak, aby pomohly v reálné praxi posoudit důvěryhodnost a původ dat. Jedním ze způsobů, jak zmást tyto metody, je po dodělení falzifikátu obrázků zobrazit na monitoru, a poté ho znovu vyfotografovat. Na detekci takto přefocených fotografií se zaměřuje druhý přiložený článek. Ten se snaží v obraze najít periodicity způsobené zobrazovací mřížkou LCD monitorů. Třetí konferenční článek řeší problém stanovení počtu referenčních obrázků k vytvoření vysoce kvalitního fingerprintu senzoru. Čtvrtý článek přináší nový postup při detekci copy-move v obraze. Nejpoužívanější techniky padělání, kdy se část obrázku zkopíruje a přenesení do jiné oblasti. Takto je možné zcela změnit informaci, kterou nese daný obrázek. Další dva články se soustředí na analýzu medicínských obrazových dat. Pátý článek představuje novou metodu zobrazování krvavých skvrn v tenkém střevě a poslední článek, popisuje detekci příznaků kmitání hlasivek z videokymografických snímků.

## ABSTRACT

---

This dissertation deals with analysis of forensic and medical image data, two fields that use advanced methods in *Image Processing*. The Thesis introduces the issue first and then describes the six attached papers with the author's contribution. The first paper presents several methods of forensic analysis of images and videos. These were designed to help in real-life practice to assess the credibility and origin of the data. One way to confuse these methods is to display the forged image on the monitor and then take another photo of it. The second paper focuses on detecting such photos. It tries to find the periodicity caused by the LCD monitor grid in the image. The third conference paper solves the problem of determining the number of reference images to construct a high-performance fingerprint sensor. The fourth paper introduces a new procedure for detecting copy-move in pictures. The most common technique of counterfeiting is to copy part of the image and transfer it to another area. In this manner, it is possible to completely change the information that the image carries. The other two papers focus on the analysis of medical image data. The fifth paper introduces a new way of displaying bleeding spots in the small intestine, and the last paper describes the detection of vibration features of the vocal cords from videokymographic images.

## PUBLICATIONS

---

### INTERNATIONAL JOURNALS WITH IMPACT FACTOR

- A. Novozámský and M. Šorel. "Copy-move forgery detection using JPEG compression model." In: *Submitted to Forensic Science International* (2017)
- A. Novozámský et al. "Automatic blood detection in capsule endoscopy video." In: *Journal of Biomedical Optics* 21.12 (2016), p. 126007. DOI: [10.1117/1.JBO.21.12.126007](https://doi.org/10.1117/1.JBO.21.12.126007)
- J. Kamenický et al. "PIZZARO: Forensic analysis and restoration of image and video data." In: *Forensic Science International* 264 (2016). Special Issue on the 7th European Academy of Forensic Science Conference, pp. 153–166. ISSN: 0379-0738. DOI: [10.1016/j.forsciint.2016.04.027](https://doi.org/10.1016/j.forsciint.2016.04.027)
- B. Mahdian, A. Novozámský, and S. Saic. "Detecting Cyclostationarity in Re-Captured LCD Screens." In: *Journal of Forensic Research* 6.4 (2015), pp. 1–6. ISSN: 2157-7145. DOI: [10.4172/2157-7145.1000294](https://doi.org/10.4172/2157-7145.1000294)

*Selected papers are  
reprinted at the end  
in their original  
form.*

### INTERNATIONAL CONFERENCE PROCEEDINGS INDEXED BY SCOPUS AND WEB OF SCIENCE

- B. Mahdian, A. Novozámský, and S. Saic. "Identification of aliasing-based patterns in re-captured LCD screens." In: *2015 IEEE International Conference on Image Processing (ICIP)*. Québec City, Canada, Sept. 2015, pp. 616–620. DOI: [10.1109/ICIP.2015.7350872](https://doi.org/10.1109/ICIP.2015.7350872)
- A. Novozámský et al. "Image analysis of videokymographic data." In: *2015 IEEE International Conference on Image Processing (ICIP)*. Québec City, Canada, Sept. 2015, pp. 78–82. DOI: [10.1109/ICIP.2015.7350763](https://doi.org/10.1109/ICIP.2015.7350763)
- B. Mahdian, A. Novozámský, and S. Saic. "Determination of Stop-Criterion for Incremental Methods Constructing Camera Sensor Fingerprint." In: *Digital-Forensics and Watermarking: 13th International Workshop, IWDW 2014*. Ed. by Y.-Q. Shi et al. Taipei, Taiwan: Springer International Publishing, Oct. 2015, pp. 47–59. ISBN: 978-3-319-19321-2. DOI: [10.1007/978-3-319-19321-2\\_4](https://doi.org/10.1007/978-3-319-19321-2_4)

#### OTHER INTERNATIONAL CONFERENCES

- B. Zitová et al. "Imaging device identification and detection of image tampering." In: *Abstract book of 7th European Academy of Forensic Science Conference*. Prague, Czechia, Sept. 2015
- A. Novozámský et al. "Computer-Assisted Evaluation of Videokymographic Data." In: *EFMI STC Prague Data and Knowledge for Medical Decision Support*. Prague, Czechia, Apr. 2013
- A. Novozámský et al. "VKFD: Computerized Analysis of Videokymographic Data." In: *10. PEVOC - Pan European Voice Conference*. Prague, Czechia, Aug. 2013
- J. Sedlář et al. "Measurement of Vocal Fold Features in Videokymography Images." In: *BioImage Informatics*. Dresden, Germany, Sept. 2012

#### INTERNATIONAL SUMMER SCHOOL

- A. Novozámský, B. Mahdian, and S. Saic. "Source camera identification invariant to zoom." In: *International Computer Vision Summer School 2012*. Sicily, Italy, July 2012. URL: <http://iplab.dmi.unict.it/icvss2012/posters.html>

#### NATIONAL WORKSHOP

- A. Novozámský. "Rima Glottidis Segmentation by Thresholding Using Graph Cuts." In: *Doktorandské dny 2012 sborník workshopu doktorandů FJFI oboru Matematické inženýrství*. Prague, Czechia, Nov. 2012, pp. 177–185. URL: <http://kmwww.fjfi.cvut.cz/ddny/historie/12-sbornik.pdf>
- A. Novozámský. "Source Camera Identification Based on PRNU Invariant to Zoom." In: *Doktorandské dny 2011 sborník workshopu doktorandů FJFI oboru Matematické inženýrství*. Prague, Czechia, Nov. 2011, pp. 163–173. URL: <http://kmwww.fjfi.cvut.cz/ddny/historie/11-sbornik.pdf>

## ACKNOWLEDGMENTS

---

My work at the field of science would have not been possible without two beautiful and great women in my life. The love of my life *Eva* and my mom *Božena*. I am so blessed to be your husband and son. Thanks for always being here for me during the good and the bad times as well. Thank you for your love, care, prayers and support. The same thanks go to my whole family, who have encouraged me my whole life.

All results described in this Thesis were accomplished with the help and support of my associates from the Institute of Information Theory and Automation which falls under the Czech Academy of Sciences.

I would like to thank my supervisor *Stanislav Saic* and all of my unofficial mentors *Barbara Zitová*, *Filip Šroubek*, *Michal Šorel*, *Babak Mahdian*, and *Jan Flusser*. Thank you for your time, patience, ideas and comments during this long period of my life.

Special thanks go to *Jan Kamenický* for his guidance of my programming skills.

I am very grateful to my two friends *Aimee Williams* and *Hana Kucová* for their assistance in proofreading of the English manuscript.

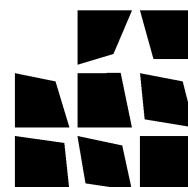
And last, but not least, I would like to thank the entire Department of Image Processing for the great working environment where I always felt very comfortable and supported. I will always gladly come back.

### My PhD studies were supported by these grants:

- TAČR - VG20102013064 - *Tools for imaging device identification, authentication, and image reconstruction*
- TAČR - VF20102012010 - *Development of methods for image analysis of photographs in digital and analog form for data authentication*
- GAČR - GA13-28462S - *Analysis of the authenticity of digital video*
- TAČR - TA04010877 - *Automatic evaluation of videokymographic recordings for early diagnosis and prevention of vocal fold tumors*
- IGA MZ ČR NT/13477-4/2012 - *Diagnostika poškození tenkého střeva nesteroidními antiflogistiky pomocí kapslové endoskopie*



Institute of  
Information Theory  
and Automation



Department of  
Image Processing



Technology Agency  
of the Czech  
Republic



Grant Agency of the  
Czech Republic



Czech Health  
Research Council



## CONTENTS

---

PREFACE      xxi

ABOUT THE AUTHOR      xxi

### I    A BRIEF INTRODUCTION      1

1    DIGITAL IMAGE PROCESSING      3

2    IMAGE FORENSICS      5

2.1    Re-capturing LCD screen      6

2.2    Source camera identification based on photo response  
nonuniformity noise      8

2.3    Copy-move forgery detection      12

3    MEDICAL IMAGING      15

3.1    Wireless capsule endoscopy      17

3.2    Videokymography      18

4    REAL-LIFE EXAMPLE - HOW CAN MEDICAL AND FOREN-  
SIC IMAGING OVERLAP      21

### II    MAIN PART OF THE THESIS      25

5    GOALS      27

6    STRUCTURE OF THE THESIS      29

7    PIZZARO: FORENSIC ANALYSIS AND RESTORATION OF  
IMAGE AND VIDEO DATA      31

8    DETECTING CYCLOSTATIONARITY IN RE-CAPTURED LCD  
SCREENS      33

9    DETERMINATION OF STOP-CRITERION FOR CONSTRUC-  
TING SENSOR FINGERPRINT      35

10    COPY-MOVE FORGERY DETECTION USING JPEG COMPRES-  
SION MODEL      37

11    AUTOMATIC BLOOD DETECTION IN CAPSULE ENDOSCOPY  
VIDEO      39

12    IMAGE ANALYSIS OF VIDEOKYMOGRAPHIC DATA      41

13    MAIN CONTRIBUTION OF THE THESIS      43

BIBLIOGRAPHY      45

### III    ATTACHED ARTICLES      59

## LIST OF FIGURES

---

Figure 1	One picture	3	
Figure 2	The oldest camera photograph	5	
Figure 3	Transverse slice through the brain	5	
Figure 4	The process of re-capturing	6	
Figure 5	Copy-move examples	7	
Figure 7	An assignment of the sensor	9	
Figure 6	Digital camera system	9	
Figure 8	The copy-move process	12	
Figure 9	Copy-move examples	12	
Figure 10	Iran missile test	13	
Figure 11	The hand of Roentgen's wife	15	
Figure 12	The first clinical CT scan	15	
Figure 13	Transverse brain slice	15	
Figure 14	Capsule endoscopy camera	17	
Figure 15	Frames of the small bowel	17	
Figure 16	The schema of a capsule endoscopy camera	17	
Figure 17	PillCam™ ESO	18	
Figure 18	The schema of VKG examination	19	
Figure 19	The output of VKG camera	19	
Figure 20	VKG frame selection 1	20	
Figure 21	VKG frame selection 2	20	
Figure 22	Root channel treatment	21	
Figure 24	Before filling	21	
Figure 25	After filling	21	
Figure 23	Analyzed X-ray images	21	
Figure 27	Revealed edge	22	
Figure 26	Intra-oral X-ray device	22	
Figure 28	Error Level Analysis	22	
Figure 29	Three modules of PIZZARO	32	
Figure 30	Natural pictures	33	
Figure 31	Re-captured pictures	34	
Figure 32	Digital camera system	35	
Figure 33	Copy-move forgery detection	37	
Figure 34	Frames without blood	39	
Figure 35	Frames with blood	40	
Figure 36	Analyzed frames of VKG	41	

## ACRONYMS

---

3D-MI	- 3D Medical Imaging
AR	- Augmented Reality
CFA	- Color Filter Array
CMOS	- Complementary Metal-Oxide-Semiconductor
CT	- Computed Tomography
DFT	- Discrete Fourier Transform
DPI	- Dots Per Inch
ELA	- Error Level Analysis
FPN	- Fixed Pattern Noise
GI	- Gastrointestinal
GUI	- Graphical User Interface
HDRI	- High-Dynamic-Range Imaging
HSVE	- High-Speed Videoendoscopy
IF	- Image Forensics
JPEG	- Joint Photographic Experts Group
LBP	- Local Binary Pattern
LBPV	- Local Binary Pattern Variance
LCD	- Liquid-Crystal Display
LED	- Light-Emitting Diode
MI	- Medical Imaging
MRI	- Magnetic Resonance Imaging
NSAID	- Nonsteroidal Anti-Inflammatory Drugs
PET	- Positron Emission Tomography
PNU	- Pixel Nonuniformity
PRNU	- Photo Response Nonuniformity
SONAR	- Sound Navigation and Ranging

SVM	- Support Vector Machines
UTIA	- Institute of Information Theory and Automation
VKG	- Videokymography
VKFD	- Videokymography Feature Detection Software
WCE	- Wireless Capsule Endoscopy

## PREFACE

---

Author's PhD studies were a great occasion to study many diverse topics in *Image Processing*, mainly *Image Forensics* and *Medical Imaging*. There have been many approaches and results, successes and failures, and some of them were presented in journals and at conferences. The key papers of his research process were selected to form this Thesis, which is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the Czech Technical University of Prague.

The rest of the Thesis is organized as follows: First, we give a brief introduction of *Image Processing* in Chapter 1, where we demonstrate the connection of two studied areas. Chapter 2 describes three methods of *Image Forensics* and Chapter 3 details two techniques of *Medical Imaging*. These methods and techniques are used in attached papers. Chapter 4 shows the real-life example of forensic analysis how can medical and forensic imaging overlap. The goals of the Thesis are concretized in Chapter 5 and the structure of the main part of the Thesis is described in Chapter 6. Then we introduce the attached papers in Chapters 7 - 12. Finally, Chapter 13 summarizes the contribution of the Thesis. At the end after bibliography, we attach six selected papers.

## ABOUT THE AUTHOR

---

*Adam Novozámský* received his MSc degree in informatics from the Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University, Prague, in 2010. He is currently pursuing his PhD with the Institute of Information Theory and Automation Cooperating Institute of Czech Technical University, Prague. His research interests include *Medical Imaging*, *Image Forensics*, and *Image Segmentation*.





Part I

A BRIEF INTRODUCTION



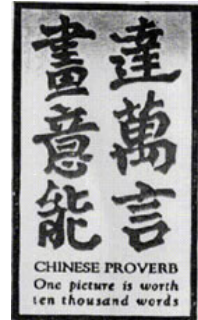
Out of all the five senses, our sight seems to be the most important. Sight and vision help people to connect with their surroundings. People like to express themselves through pictures and approximately 65% of the general population are visual learners [13]. Nowadays we live in the world with technologies such as Augmented Reality (AR), High-Dynamic-Range Imaging (HDRI), or 3D Medical Imaging (3D-MI). We meet images or videos everywhere around us, on the Internet, in newspapers and magazines, on television, at the doctor's office, or in private family albums. No wonder that *Image Processing* can be found in thousands of scientific, consumer, industrial, and artistic applications. This amazing diversity of applications, modalities, and image types makes *Image Processing* such an interesting topic to study, where every project needs a genuine, innovative approach. The author of this Thesis had a great opportunity to study many specialized topics in the fields of *Image Forensics* and *Medical Imaging*.

The various fields of *Image Processing* are intertwined and share many approaches, techniques, and mathematical background. This also applies to both selected topics. There is a strong link between *Medical Imaging* (MI) and *Image Forensics* (IF). They not only use similar methods and algorithms but they meet each other in applications. Medical images often help in crime investigation, are used as evidence at a court and may be a subject of forgery. Their forensic analysis is required in these situations. On the other hand, method originally developed for forensic purposes (such as specialized super-resolution algorithms) have found numerous applications in medical imaging.

Medical images have been used in criminology since the beginning. Only a matter of weeks after Roentgen's discovery, the forensic use of X-rays helped to convict an American attempted murderer [2]. The same holds for other modalities being used not only for diagnosing diseases but also as an evidence in the court of law; such as Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) [71]. Nowadays, there are plenty of articles and some journals dedicated solely to this area.<sup>1</sup>

Post-mortem imaging or virtual autopsy are good examples of the application of imaging in forensics. Both are an alternative to the traditional invasive autopsy, namely beneficial for relatives of trauma

<sup>1</sup> For example *Journal of Forensic Radiology and Imaging* published by Elsevier and affiliated with the International Society of Forensic Radiology and Imaging [128] and the International Association of Forensic Radiographers [62]



**Figure 1:**  
Chinese proverb.  
One picture is  
worth ten thousand  
words [3].

victims. In 2015, Jalalzadeh et al. published an extensive review comparing post-mortem imaging and conventional autopsy in trauma victims [63]. They systematically searched in journal databases for studies published in years 2008-2014, where the victims were investigated by both conventional and imaging methods. They found twenty-six studies with a total number of 563 trauma victims. Their conclusion was that, with limitations, the imaging approach could be used as an alternative in situations where conventional autopsy is rejected or unavailable.

*Image Forensics (IF)* is a branch of *Forensic Science* which focuses on image authentication and identification of the camera device that captured the image. The first manipulated photographs appeared in the 19th century, when absent individuals were inserted into portraits and heads of politicians were placed on other people's bodies [61], see the introduction in Chapter 2. However, IF in the way we know it today originated in the second half of the 20th century along with PC development, especially with the advancement of image editing tools. These days, we have a lot of sophisticated techniques dealing with image tampering recognition. Some of them were originally developed in MI, as well as in other areas of *Image Processing*.

We see this overlap of both studied branches of *Image Processing* also in some problems being solved at our Department of Image Processing, Institute of Information Theory and Automation (UTIA). Our department is a certified forensic institution in computer science, with a special focus on assessing the authenticity of digital images and videos. The medical images are among those being analyzed to a significant extent. A typical scenario is a lawsuit between a doctor and a patient, where the evidence is, for example, a patient's X-ray picture before surgery. The forger can be on both sides and even a small change may justify, confirm or refute the procedure according to the law of the art (*lege artis*). Chapter 4 shows one such case.

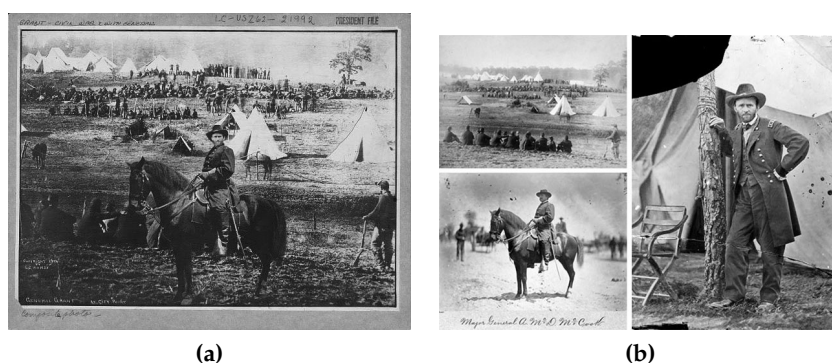
IF is very closely related to the history of analog photography. One of the first tampered images originated a few decades after Niépce created the first photographs in 1826, see Fig. 2. A nice example of the analog image splicing was created sometime around the year 1864, during the American Civil War. It is a photograph of General Ulysses S. Grant in front of his troops, Fig. 3a. However, detective work of the Library of Congress successfully proved that this image is a collage of three photos, in Fig. 3b.

Tampering with images is easy with today's software tools for *Image Processing*. A forger can add or remove important information which completely changes the message of a processed photo. The integrity of visual data is important for the credibility of news media and especially when used as an evidence in court or during criminal investigations. For this reason, we have observed a dynamic development in this research area in recent years. Over time, two branches of forensic analysis of digital images and videos have become established as essential. One of them is integrity verification (authenticity analysis), which determine if material has been added, altered, deleted or changed in the image. The second one is image ballistics (source device verification), which identifies the source device of acquisition process of the image.

Although past research in these areas has mainly focused on data hiding and digital watermarking approaches, today there is a growing alternative approach called the passive one, which does not need to embed any secondary data into the image [91]. Methods published in this research area have, for example, attempted to detect image splicing [20], traces of inconsistencies in color filter array interpolation [38], traces of geometric transformations [18], cloning [43], compu-



**Figure 2:**  
The oldest surviving camera photograph: "View from the Window at Le Gras", Joseph Nicéphore Niépce, ca. 1826 [40].

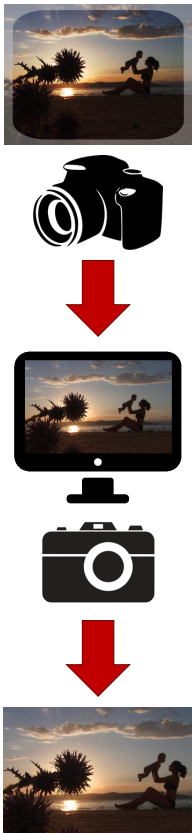


**Figure 3:** One of the first tampered images (a) and its originals (b) [112].

ter graphics generated photos [28], JPEG compression inconsistencies [90], file structure inconsistencies [35], etc. Typically, these methods are based on the fact that digital image editing brings specific detectable statistical or geometrical changes into the image. Others are looking for a distortion in the reality of the image scene.

The contribution of our department to this area of research is covered by the paper [66]. The author of this Thesis dealt primarily during the PhD studies with the re-capturing LCD screen [87], source camera identification based on Photo Response Nonuniformity (PRNU) noise [88], and copy-move forgery detection [100]. The following three sub-chapters consist of a brief introduction to these disciplines.

## 2.1 RE-CAPTURING LCD SCREEN

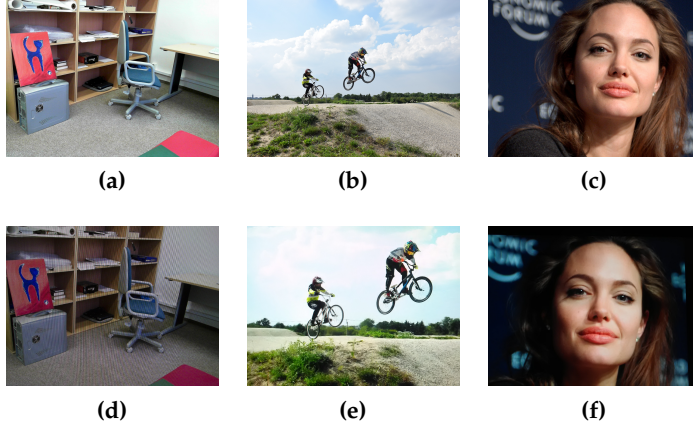


**Figure 4:** The process of re-capturing LCD screen.

Today, there is a widespread availability of high quality LCD that can be used for re-capturing displayed images by using a digital camera or smartphone. The re-capturing process is in Fig. 4 and the images from re-capturing are shown in Fig. 5. Images 5a, 5b, and 5c represent three original versions. Images 5d, 5e, and 5f demonstrate corresponding re-captured images obtained from displayed images on LCD screens. Please note that there are different qualities of re-captured images. The image 5d is poor and exhibits visible traces of re-capturing easily detectable by the human eye, 5e is medium re-capturing quality and harder to recognize for the human eye. The image shown in 5f is finely re-captured and traces of re-capturing are not visible to the human eye.

Detecting re-captured images has an important role in several areas. One way of validating images is by searching for traces of digital manipulation. Most pixel-based methods for detecting image manipulation are based on searching for inconsistencies among pixels [36]. They assume that the digitally manipulated part of the image might exhibit different JPEG and quantization compression artifacts [37] compared to the genuine part of the image, etc. Typical additional features used to detect digital manipulations are sensor noise properties in various parts of the image [83] or estimating intra/inter bands correlations among pixels [45]. All these methods can be spoofed by displaying the image on an LCD screen and re-capturing the photo. The secondary (re-captured) image will behave as an original digital image and its pixels will not exhibit any statistical inconsistency.

In addition to the impact of re-capturing mentioned, it is important to note that face recognition has become a standard biometric authentication technique in multiple applications. In contrast to other biometric authentication techniques such as fingerprints [52, 69], iris recognition [86, 127], etc., face recognition [31, 76, 122, 138] is well compatible with consumer webcams and cameras that laptops and



**Figure 5:** Shown are three examples from LCD re-capturing.

mobile phones are today equipped with. Face recognition does not require any additional special sensor. This is an exceptional advantage that made face recognition a widely used method. On the other hand, there are serious concerns about possible face spoof attacks (i.e., when a person tries to masquerade as someone else). Here, one can easily print an image of a person from a Facebook account, and spoof access systems based on face recognition [9, 81].

To partially overcome the problem, face recognition based on unlocking and authentication solutions are often enriched by liveness detection features (e.g., often based on blinking [106]). However, such solutions are often vulnerable in scenarios where spoofing attacks are performed by displaying an image or video on LCD screens.

Another impact of the ability to detect re-captured images and videos is to help detect video piracy, where movie videos acquired in cinemas can be automatically detected on sites such as YouTube.com, etc.

Although most of the so-far published methods of detection of re-capturing LCD are focused on a specific domain (e.g. face spoofing), a few papers have also been published which propose methods for general image re-capturing. For example, LCD re-capturing has been addressed in Thongkamwitoon et al. [131, 132]. Here, authors have proposed an LCD re-capture detection algorithm based on learned edge blurriness and distortion. Two sets of dictionaries are trained using the K-singular value decomposition approach from the line spread profiles of selected edges from single captured and re-captured images. A support vector machine classifier is then built using dictionary approximation errors and the mean edge spread width from the training images. Their method achieves a performance rate that exceeds 99% for re-captured images and 94% for single captured images. They have experimentally shown a powerful generalization and that the proposed algorithm is robust to a wide range

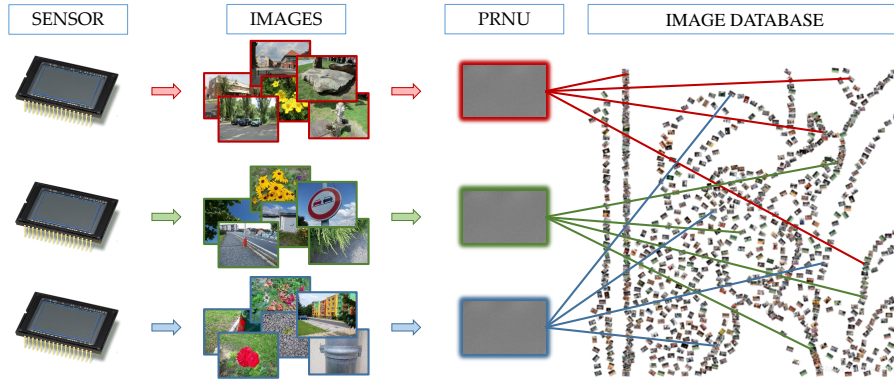
of images and can be applied successfully to a wide range of capture devices. Using another method, our team under the leadership of Mahdian [89] has shown that re-captured images and videos from LCD screens exhibit detectable periodic patterns that are caused by the regular sampling grid of the LCD monitor and aliasing. We proposed a method to detect if a signal has a frequency spectrum correlated with a shifted version of itself. This method uses the theory of cyclostationarity, a special class of signals which exhibit periodicity in their statistics [46]. Bestagini et al. [8] have proposed a detector based on the analysis of a characteristic ghosting artifact left during the re-capturing process. In particular, since ghosting can be modeled as an effect of the filtering operation, our novel approach exploits the motion between frames to retrieve information about the ghosting filter (i.e., the shape), and verifies if such a filter may have been used on the frames. If the answer is positive the frames are considered re-captured. Cao and Kot [14] have proposed a set of statistical features, which capture the common anomalies introduced in the camera re-capturing process on LCD screens. These features include the texture patterns, the loss-of-fine-details characteristics and the color anomalies introduced in the image re-capturing process. Scarzanella and Dragotti [134] have proposed using scene jitter as a feature for classification of re-captured videos. Kose and Dugelay [68] proposed a method based on analysis of contrast and texture characteristics of captured and re-captured images. Here, specifically, rotation invariant Local Binary Pattern Variance (LBPV) based method is used. Määttä et al. [85] proposed a method to identify face spoofing by analyzing texture of the facial images using multi-scale Local Binary Pattern (LBP) [141]. Pereira et al. [111] proposed a method to detect face spoofing using the spatiotemporal (dynamic texture) extensions of the highly popular LBP operator.

We can notice that most of the methods use hand-engineered features such as LBP, noise estimators, etc. Also most of them are tailored for a specific and narrow domain.

## 2.2 SOURCE CAMERA IDENTIFICATION BASED ON PHOTO RESPONSE NONUNIFORMITY NOISE

Source camera identification deals with image (video) ballistics which does address the problem of linking digital images (videos) under investigation to the exact source imaging device that has been used to capture them. Since image ballistics makes it possible to differentiate between source camera devices of the same make and model, it has become useful especially in the digital forensic area.

In the captured images, *Pattern Noise* consists of two components, called Fixed Pattern Noise (FPN) and Photo Response Nonuniformity (PRNU) noise. FPN is caused by dark currents, which is independent



**Figure 7:** An assignment of the sensor to images in the database.

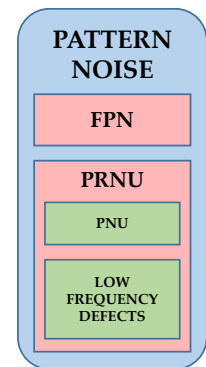
of the signal. It is an additive noise depending on the exposure and the temperature. Some middle and high-end consumer cameras can suppress this noise automatically [84]. The dominant part of *Pattern Noise* is **PRNU**, which consists of Pixel Nonuniformity (**PNU**) and low frequency component.

The **PNU** originates in varying pixel dimensions and inhomogeneities in the photo sensor. These imperfections lead to variations of the sensitivity of pixels in the final output. This part of noise is multiplicative and its values increase with the signal level (it is more visible in pixels represented in light scenes). Moreover, it is not dependent on temperature and it seems to be stable over time. The low frequency component contributes only negligibly to the **PRNU** and is caused by reflection on dust particles, optical surfaces, and the sensor itself.

Fridrich et al. [19, 42, 83, 84] have shown that it is possible to perform image ballistic analysis by using **PRNU** as a unique identification fingerprint of the camera sensor. To identify the exact source camera for a given image, see Fig. 7, **PRNU** is considered to be a spread spectrum watermark, whose presence in the image is examined by using a correlation detector. The process of source camera identification is described in greater depth in our attached paper [88].

Since the first article on the subject appeared, we have registered a rapidly growing trend in the number of new publications using the sensor noise as the camera fingerprint. By going over them, it is easy to notice that most of them are using the core idea published by Lukáš et al. [84]. Published methods have often brought extensions which suggest applications for new uses, or empirically-based improvements to the basic idea.

Application of the methods in real-life environment outside the scientific labs easily reveal that this field is not yet fully developed. Uncontrolled environments often cause unexpected and unreliable results because of high number of variations and complexities of hardware and software components that imaging devices are equipped with. Moreover, there are still a number of essential topics that have



**Figure 6:** Pattern noise of imaging sensors; inspired by Fig. 1 in [84].

remained unanswered by scientists so far. On the other hand, the growing number of scientific contributions to the field signifies that we can expect remarkable enhancements and progress in the coming years.

In recent years, we have seen papers that have focused on the denoising filter that is a common part of PRNU estimation. For example, Zhang [140] has proposed a new camera identification method that uses a curvelet-based filter to extract noise features in heavily textured images. By modeling a sensor pattern noise as a white noise, Wu et al. [137] have proposed a method that extracts sensor pattern noise directly from the spatial domain using an adaptive Wiener filter instead of using wavelet high frequency coefficients used in most of the other methods. Amerini et al. [4] have empirically analyzed the performance of different denoising filters for PRNU analysis. Similarly, Conotter and Boato [23] have analyzed performance comparison of different denoising filters for source identification purposes and have discussed the topic with respect to state-of-the-art denoising algorithms. Knight et al. [67] have proposed modification of the denoising filter in PRNU estimation, allowing it to be used for raw sensor data. Cortiana et al. [26] have also made a performance comparison of different denoising filters for PRNU identification. Cooper [24] has demonstrated that an adaptive spatial filter coupled with a median filter can outperform methods using wavelet based denoising.

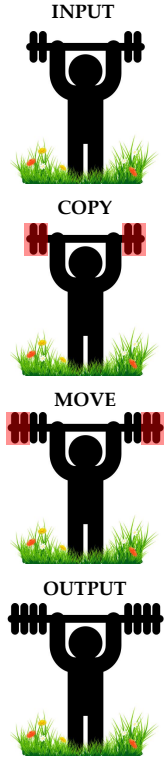
Another group of methods has analyzed image ballistics in respect to image geometrical degradations. For example, Goljan and Fridrich [49] have extended their basic camera identification technology to a more general setting when the image under investigation has been simultaneously cropped and scaled. Goljan and Fridrich also have proposed selecting the signal subset based on the outlier values in the sensor fingerprint [50]. The advantage is that the test statistic for such a subset decreases substantially slower with decreasing signal length, thus one obtains a much better trade-off between the gain in speed and the detection accuracy. Rosenfeld et al. [118] have analyzed the robustness of PRNU-based camera identification in cases of presence of common *Image Processing* operations such as scaling/-cropping, geometric distortion, etc.

There is also a group of publications aiming to enhance results by better understanding of image content, CFA, or by having different matching technique for PRNU signals. For example, Chan et al. [17] have tried to reduce the image content effect in classification by using a correlation predictor that is trained based on different types of image features. Matsushita and Kitazawa [93] have improved PRNU camera identification results by selecting appropriate pixels used for correlation according to the texture complexity of image. Li et al. [75, 77] have extracted PRNU by exploiting the difference between the physical and artificial color components of the photos

taken by digital cameras that use a color filter array for interpolating artificial components from physical ones. Orozco et al. [25, 105] have proposed jointly employing sensor imperfections and wavelet transform as features for forensic analysis. Their approach utilizes the optimal parameter search from SVM for prediction and classification. Bayram et al. [7] have proposed representing sensor fingerprints in binary-quantized form. They have shown that the reduction in matching accuracy due to quantization is insignificant as compared to conventional approaches. McCloskey [94] has proposed a confidence weighting scheme to address the problem of extracting a fingerprint from videos where high-frequency content (e.g., edges) persist at a given image location. Li [73] has envisaged the hypothesis that the stronger a component of the PRNU is the less trustworthy. Hence, he has assigned greater weighting to the smaller PRNU components. Liu et al. have [79] proposed to extract from the noise residual the significant regions with higher signal quality and discard those regions heavily deteriorated by irrelevant noises. Hu et al. [56] have tried to build the camera fingerprint based on three color bands of an image. Liu et al. [80] have formulated image ballistics as a graph partitioning problem so a set of images can be classified according to their source cameras in an entirely blind way. Hu et al. [57] have used matching signs between query and reference fingerprints to perform a fast source camera identification. Cattaneo et al. [16] have analyzed a Hadoop distributed approach to solve the camera identification problem.

Methods for clustering images based on PRNU have also been published. For example, Baar et al. [5] have investigated methods for grouping single images based on PRNU in large databases. Bloy [11] has proposed a simple clustering method to cluster images of the same PRNU into groups. Bayram et al. [6] have analyzed the efficiency of the sensor fingerprint based source device identification technique in large databases.

Reliability of PRNU-based fingerprints have also been discussed from a different perspectives. For example, Goljan et al. [51] have conducted a large scale test of camera identification from sensor fingerprints. The experiments have helped to establish upper bounds on error rates of their camera identification method. Cattaneo et al. [15] have analyzed the effect of JPEG compression on methods dealing with PRNU. Li et al. [74] have discussed the problem that image sensor noise can easily be manipulated. Dirik and Karakucuk [30] have introduced an image source anonymization method against PRNU-based source camera identification. Steinebach et al. [124] have shown how camera forensics can be attacked and introduce a potential countermeasure against these attacks.



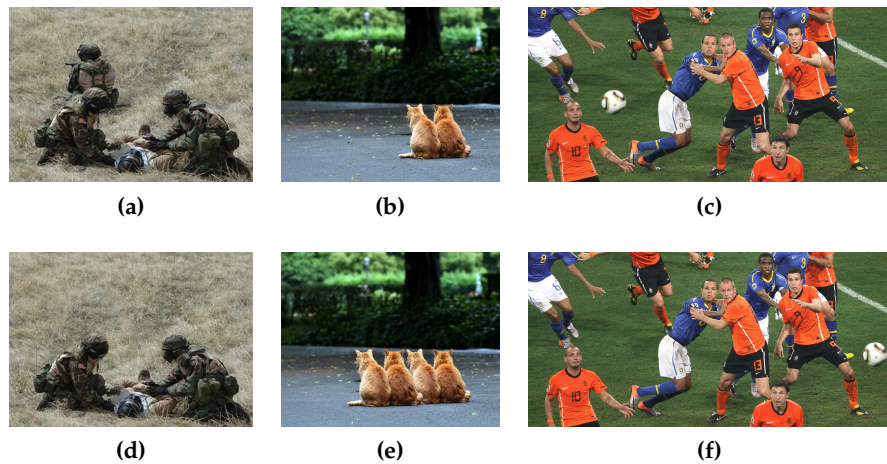
**Figure 8:**  
The copy-move  
process in a image.

### 2.3 COPY-MOVE FORGERY DETECTION

A common and simple way to manipulate image content is the copy-move forgery, when someone copies an object and pastes it in another location of the same image, see Fig. 8. Anyone who has ever edited a picture with the cloning tool in Photoshop knows how fast and easy it is. The motivation for using this tool can be covering someone or something in the picture (Fig. 9d), replicating (Fig. 9e), highlighting, clipping, or moving the object content (Fig. 9f).

It is quite easy to uncover the fake images when one also sees the original like in Fig. 9. But if we see only the modified image, it may become impossible for the human eye. One infamous copy-move attack was the Iran missile test, shown in Fig. 10. On July 9, 2008 the Iran’s Revolutionary Guard reported a successful missile launch confirmed with the picture 10a in its press. It was immediately reprinted on the front pages of The New York Times, The Los Angeles Times, The Chicago Tribune and several other newspapers as well as other major news web sites. One day later, The Associated Press received another image, Fig. 10c, which appeared to be taken from the same vantage point at almost the same time [60].

Thanks to these cases, new approaches were developed to authenticate the copy-move in digital images. And in recent years, this research topic has become one of the most actively examined in blind *Image Forensics* and a considerable number of different methods have been proposed. We can divide these algorithms into two main groups. First is the block-based and the second is keypoint-based. The description of most popular copy-move forgery detection techniques as well as their division into groups is summarized in the literature su-



**Figure 9:** Copy-move examples. The originals are on the first line (a [90], b [29], c [130]) and their edited versions on the second line.



**Figure 10:** Copy-move attack in Iran missile test: picture submitted by Iran with four missiles (a) [96], a SURF-based search on any copy of the image (b) [12], the original image (c) [96], and the funny picture (d) [27] with many missiles.

urvey in our paper [100]. For their evaluation we recommend reading the paper from Vincent Christlein et al. [21].



## MEDICAL IMAGING

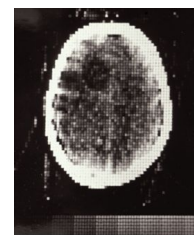
*Medical Imaging* has a long history reaching back to the end of the 19th century, when in 1895 Wilhelm Conrad Roentgen took the first medical image, the radiograph of the left hand of his wife Bertha, see Fig. 11. Only two months later he performed the first clinical use and after half a year he presented a new kind of ray [116]. His discovery had a profound impact on medicine. He was awarded the first Nobel Prize in Physics [72]. Many techniques were developed based on his approach, but the projection stayed two-dimensional for many decades until 1969.

In this year, Godfrey Hounsfield used mathematical methods for image reconstruction Allan Cormack's and produced the first Computed Tomography (CT) scanner. Two years later, he demonstrated the CT scan of animals and later also the scan of human head, as shown in Fig. 12. Cormack and Hounsfield shared the Nobel Prize in Medicine in 1979 [72].

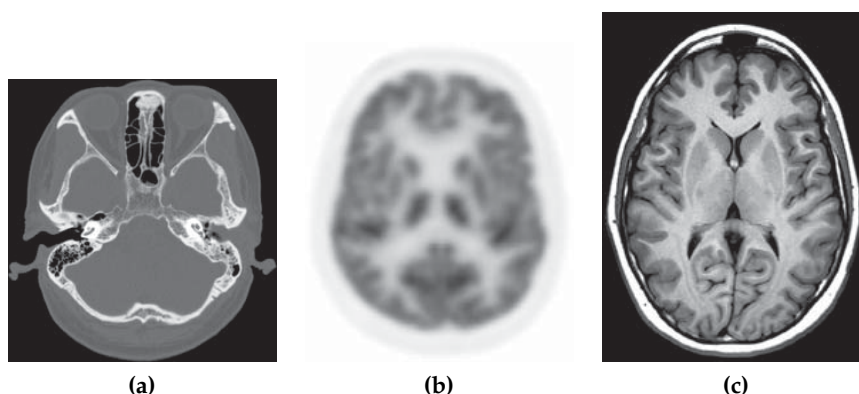
Thanks to X-rays, doctors had the non-invasive way to see the anatomical structures of patients, but only a static one. The movement and flow of substances could not be observed with methods based on X-rays. Fortunately, in 1923 the concept of using radioactive tracers to study physiology was introduced by George de Havesy [113]. Until then radionuclides were used only in cancer therapy in medicine. These radiotracers are nowadays radioactively labeled drugs in which one or more atoms have been replaced by positron emitting radioisotopes [22]. Havesy's idea led to the invention of emission and transmission tomography, which was introduced by David E. Kuhl,



**Figure 11:**  
The hand of  
Roentgen's wife  
Bertha in  
December 1895.



**Figure 12:**  
The first clinical  
CT scan, 1st  
October 1971,  
Atkinson  
Morley's  
Hospital [1].



**Figure 13:** Transverse brain slice from three different imaging modalities: CT (a), PET (b), and MRI (c) [113].

Luke Chapman and Roy Edwards in the late 1950s. And after another twenty years, the first Positron Emission Tomography (PET) scanner was developed Michel Ter-Pogossian's team [126]. The comparison of these first three imaging methods is shown in Fig. 13.

Another technique for seeing the human body on the inside uses sound. In the beginning, sound waves were used to measure distance underwater. The first pioneer of this method was John William Strutt, 3th Baron Rayleigh. He published *The Theory of Sound* [115], where he described vibrations and resonance in elastic solids and gases. The first volume appeared in 1877, followed by a second, concentrating on acoustical propagation in material media [78]. Outputs from his research were used to construct the SONAR (acronym of Sound Navigation and Ranging) in the United States by Canadian Reginald Fessenden in 1914 [39]. Similarly to nuclear medicine, ultrasonic waves were first used in the field of medicine as therapy rather than in medical imaging. The possibility of ultrasonic diagnosis based on echo-reflection methods was presented by H. Gohr and Th. Wedekind in 1940 [48], however they were not able to present the results [55]. First to show the results was Karl Theodore Dussik [32]. Ultrasound technology progressed through the 20th century to today's two-dimensional (2-D) Doppler, three-dimensional (3-D), and nonlinear imaging systems [113].

In 1952, Felix Bloch and Edward Mills Purcell were awarded the Nobel Prize in Physics [72]. They developed new ways and methods for nuclear magnetic precision measurements. That was the foundation for the discovery of Magnetic Resonance Imaging (MRI), awarded the 2003 Nobel Prize in Physiology or Medicine. The holders are Paul Lauterbur of the University of Illinois and Sir Peter Mansfield of the University of Nottingham [129].

Radiography, CT, PET, Medical Ultrasound, and MRI, these five techniques are representatives of *Medical Imaging*. The last two are non-invasive and risk-free so far as we know. The first three are considered non-invasive as well, beside some risks associated with radiation exposure. Until recently, we have had to consider imaging in the visible spectrum as invasive, carried out through a surgery or classic endoscopy. But today's progress brings miniaturization and new technologies to overcome this issue.

Over the past several years, our department has worked closely with leading medical institutions in Czechia. Among other things, this long-lasting partnerships has resulted in two medical projects on which the author of this Thesis collaborated. Both are very specialized and deal with the display of the human body in the visible spectrum. The first one is the Wireless Capsule Endoscopy (WCE) focused on imaging of the Gastrointestinal (GI) tract. The second one is the Videokymography (VKG) focused on imaging vocal cord function. The following subsections describe these imaging techniques.

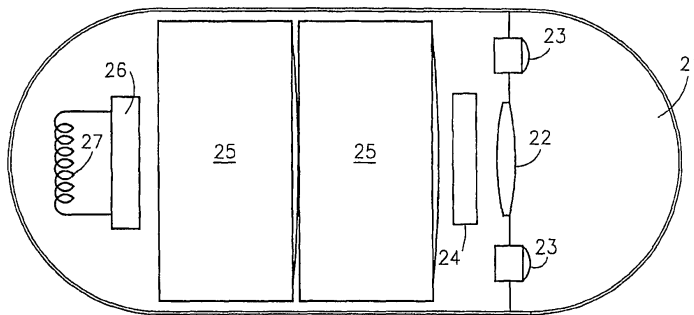
### 3.1 WIRELESS CAPSULE ENDOSCOPY

**WCE** is a non-invasive diagnostic tool very considerate to the patient. It has enabled, for the first time, a painless diagnosis inside the **GI** tract including the small bowel [95]. The pioneer company in this field of research was *Given Imaging* [59], which brought **WCE** into clinical practice in 2001 and filed a patent application in 2006 [58]. Over time, it has been successfully applied in detection of small bowel bleeding sources, Crohns disease, complications of coeliac disease, tumors and **NSAID** induced enteropathy.

The patient swallows a cylindrical plastic capsule. The capsule travels through the **GI** tract by peristaltic contractions, captures the images and transmits them in real time wirelessly to an external console worn by the patient. The images are recorded and stored in the console memory and can be uploaded to a computer for a visual inspection or automatic analysis immediately after the monitoring has been completed.

This method has expanded greatly in the past decade and new capsules have appeared. One of them is the PillCam™ ESO (Given Imaging), Fig. 17, which was developed for the imaging of the esophagus. It has a faster frame rate (18fps) and images are acquired in a shorter time; 25 minutes until the capsule automatically turned off [33].

The schema is in Fig. 16. The capsule typically consists of an optical window 16(21) and an imaging system for obtaining images. The imaging system includes a white **LED** illumination source 16(23), a **CMOS** imaging camera 16(24), and an optical system 16(22). The capsule further includes a transmitter 16(26) and an antenna 16(27) for transmitting the video signal, and a power source 16(25). The schema in Fig. 16 and its description are taken over patent [58]. The size of the capsule depends on the manufacturer and is usually about 10x25 mm. One of them is shown in Fig. 14. The first capsule from the group of Iddan was 11x30 mm [59].

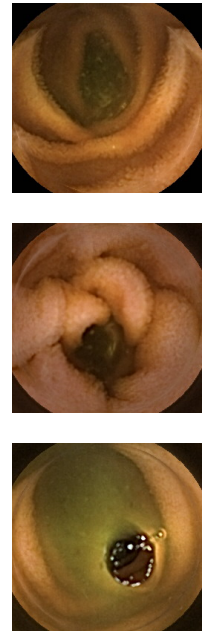


**Figure 16:** The schema of a capsule endoscopy camera. The description of the image is included in the body of the text [58].

*Nonsteroidal Anti-Inflammatory Drugs (NSAID)* is a drug class that reduces pain [10].



**Figure 14:** Capsule endoscopy camera [47].



**Figure 15:** Samples of frames of the small bowel acquired by **WCE**.

*Motility is a term used to describe the contraction of the muscles in the GI tract [136].*



**Figure 17:**  
*PillCam™ ESO*  
[47].

Internal batteries are powerful enough to supply the light source and complete the acquisition and transmission process, lasting up to 10 hours. Current capsules take frames at a rate between two and six per second on average, which results in many thousands of images; typically over 60,000. The image resolution is mostly 256x256 pixels; Fig. 15 shows the samples.

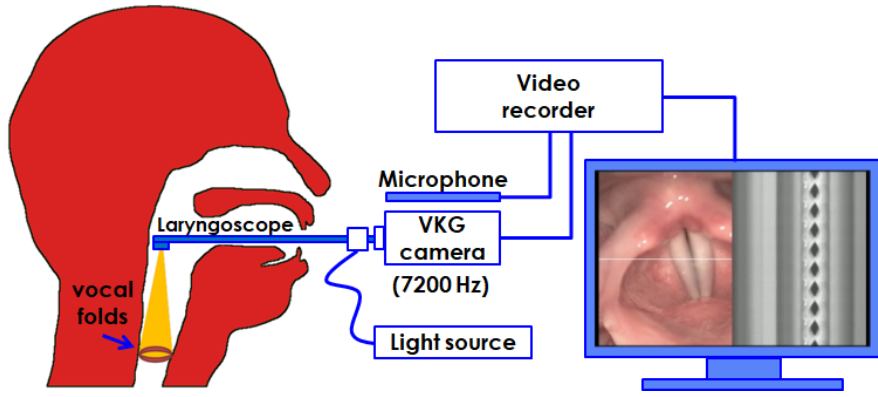
It should be noted there are patients for whom WCE is not appropriate. It is not recommended to use WCE in those with known motility disorders, swallowing disorders or dementia. On the other hand, we can label WCE as an extraordinarily safe device to be used. Over 340,000 capsules have been deployed worldwide with no reported deaths and with a few side effects, published in 2006 by Pannezio [110]. If there is suspicion of retention of the capsule in the GI tract, the check is carried out by normal X-ray and after the confirmation surgery is performed. For this reason, WCE is suitable only for patients who are medically fit for surgery [109].

The main obstacle in the massive usage of WCE is that the visual evaluation of the video is very time-consuming. It is supposed to be done by a trained endoscopist. Even if the software provided by the capsule producers offers certain supporting tools to simplify and speed up the process, it still takes at least one hour of a full concentration of the evaluator; in many cases over two hours. This led to the research of new algorithms for blood, ulcer, and tumor detection in the bowel.

Since the bleeding detection is a very frequent requirement and appears to be easily achievable, the largest group of papers has dealt with this challenge. The problem here is that the blood spots and traces do not have any typical texture and shape, and the blood color may vary in a wide range from light red through dark red to brown. This diversity in color is depending on types of disease, bleeding time [70], the position of the capsule and the surrounding conditions [107]. The approaches, adopted by various authors differ from one another only in the color space they are working in. Traditional color spaces such as RGB and HSV have been commonly used but some authors proposed also specialized color spaces. Most of the methods extract the features at the pixel level [65, 108, 121], others use image blocks [70] and some methods work with the image as a whole [139].

### 3.2 VIDEOKYMOGRAPHY

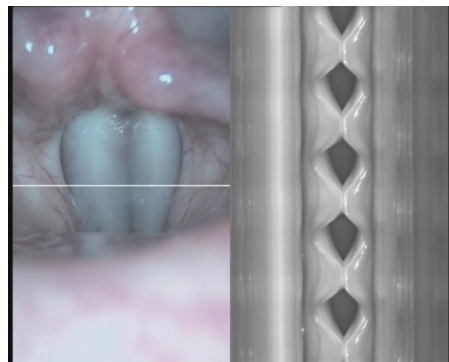
The second medical imaging in the visible spectrum, which is a part of this Thesis, is Videokymography. It is used in laryngology and phoniatrics for observation and diagnosis of vocal fold vibrations. The schema for VKG examination is shown in Fig 18. It is an original Czech-Dutch method [125], introduced in 1996 in Groningen (NL) as an alternative to videostroboscopy and High-Speed Videoen-



**Figure 18:** The schema of VKG examination.

doscopy (HSVE). Neither technique is feasible for most laboratories. The stroboscopy is a well known technique in laryngology [117] and quite simple, but the slow motion of the vocal folds is an illusion. In addition, it is only applicable to healthy vocal cords with pure periodicity. Herriott and Farnsworth presented the usage of HSVE in Journal of the Acoustical Society of America [53]. Various studies followed, which are summarized in Hirano's survey [54]. In 2008, Lohscheller et. al [82] transformed the segmented edges of vibrating vocal folds from HSVE images into a single 2-D image, denoted phonovibrogram (similar to the videokymogram). But the HSVE equipment is expensive and the analysis of such videos is time-consuming.

A VKG camera acquires brightness values only in one row but with a high-speed rate of 7200 lines/s. It can thus capture gradual movements of vocal folds, albeit in one dimension. The resulting videokymogram is a spatio-temporal image of vocal fold vibrations along a line perpendicular to the glottal dorsoventral axis over a short period of time. The advantages of VKG include low costs and a smaller amount of data to process. The final output of the video stream consists of two parts, Fig. 19. On the right side there is the videokymogram and on the left side is the normal 2-D image of the vocal folds for checking the correct location of the camera by the doctor.



**Figure 19:** The output of VKG camera by vocal fold vibration.

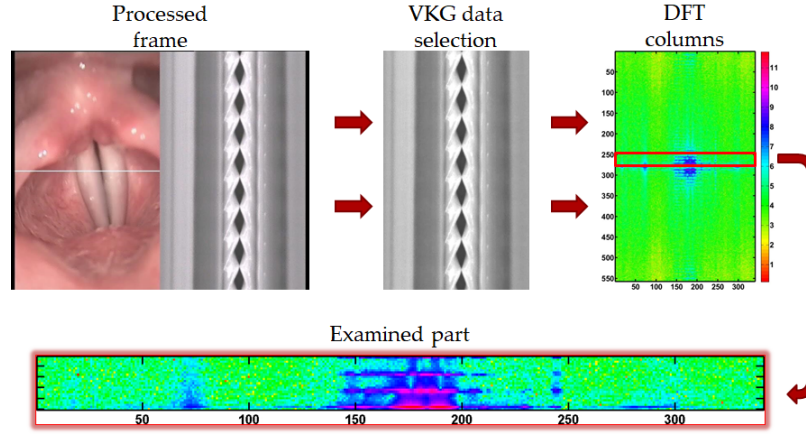


Figure 20: VKG frame selection based on DFT - with vibration.

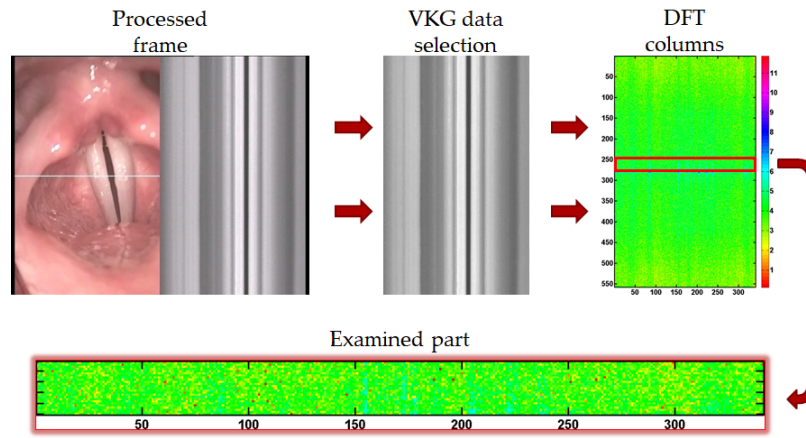


Figure 21: VKG frame selection based on DFT - no vibration.

Manual analysis of videokymograms is, however, difficult and also time-consuming, mainly due to the noise, reflections and low contrast of the images as well as high number of vocal fold parameters. The possibility of computer-aided diagnostics is thus of great interest. Although, the problem has been addressed in several papers, it has not been solved yet. In 2003, Qui et. al [114] proposed the automatic detection of some parameters of vocal folds vibration, based on an active contour model with a genetic algorithm. Manfredi et al. [92] developed an algorithm based on snake active contours in 2006. But these algorithms and others were sensitive to low contrast between vocal folds and glottal space.

Thus, the main objective of the second medical paper included in this Thesis was to develop (semi)automatic methods for the measurement of selected vocal fold features also for low contrast videokymograms. The side product was the selection of all parts of the VKG video, where the vocal fold vibrates. For this purpose we used the Discrete Fourier Transform (DFT), shown in Figures 20 and 21.

## REAL-LIFE EXAMPLE - HOW CAN MEDICAL AND FORENSIC IMAGING OVERLAP

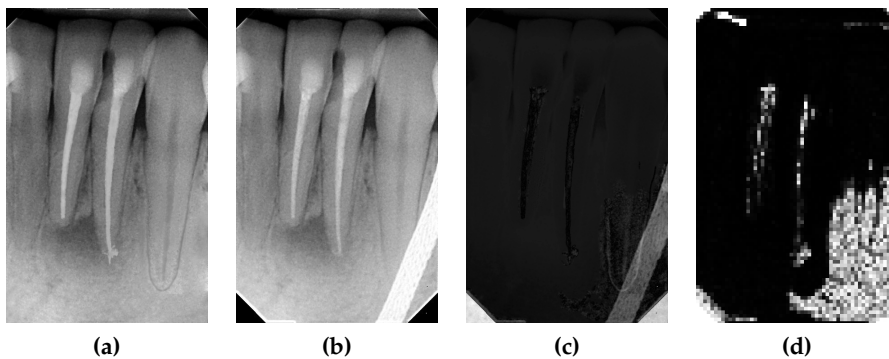
Here we demonstrate the real-life example of *Forensics imaging* and in the future, we expect an increase in the number of similar cases. We describe one of the expertise reports prepared for a trial by our Institute, also mentioned in our journal article [66]. The subject of the dispute was the X-ray image in JPEG format, Fig. 23a, which the dentist presented as proof of having correctly filled the root canal of the tooth. Correct filling means that the canal is filled throughout its length, see the procedure in Fig. 22 and the X-ray image of a proper filling in Fig. 24 and 25.

In the court file there was one more X-ray image, Fig. 23b. The dentist claimed that the image in Fig. 23a is a separate X-ray scan, taken at a later date than Fig. 23b. In addition, he declared that Fig. 23a was a scan of a paper-printed version of the X-ray scan. Our task was to perform the forensic analysis of Fig. 23a.

HERE WE MENTION SOME OF OUR OBSERVATIONS:

(I) It is almost impossible to take two X-ray images at different times from the same place. Neither the patient nor the X-ray sensor are fixed and could not be in the same position for both scans see the device setting of the intra-oral X-ray scanner in Fig. 26. This means that the angle of rotation relative to the jaw and the distance of acquisition has to be different. So we should see two different views. But here we see almost the same images (pixel by pixel).

- **Pixel levels:** If we subtract the images from each other, there is no difference in the major part of the image, Fig. 23c.



**Figure 23:** Analyzed X-ray image (a), the second image in the lawsuit (b), their pixel difference (c) and the correlation mask (d).



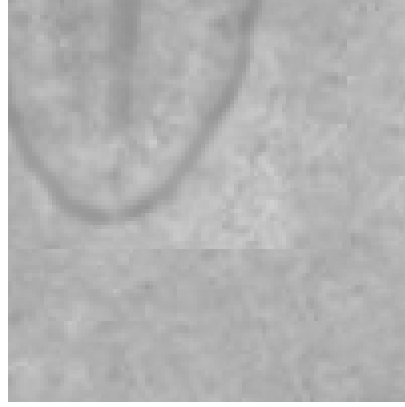
**Figure 22:**  
The process of root canal treatment [34].



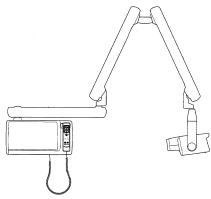
**Figure 24:**  
Before filling - The abscess present in jawbone [see red box] [119].



**Figure 25:**  
After filling - Proper root canal and complete bony healing [119].



**Figure 27:** Revealed edge from the copy-move process.



**Figure 26:**  
A side view of a  
intra-oral X-ray  
device. [64].

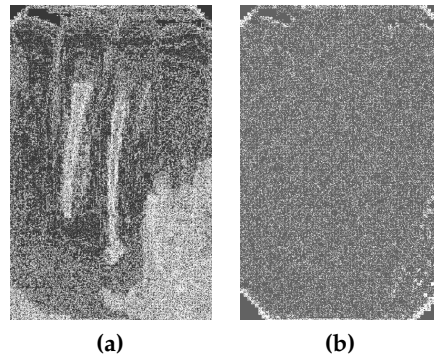
ELA reveals areas  
within an image  
that are at  
different  
compression levels.

- **Block correlation:** If we count the block correlation between images we see low values on most parts of the image and high in locations of the expected change, Fig. 23d.

( II ) The dentist claimed that Fig. 23a is the scan of the physical copy of an X-ray image. How is it possible that the scanned image has the same resolution as Fig. 23b, and the same DPI?

( III ) If we look at one of the teeth, which is partially covered by a white band in image 23b, we can see some suspicious straight lines in Fig. 23a. The zoom of this area is shown in Fig. 27. A line of mismatched gray levels like this would never appear on a natural X-ray. We would only see this kind of mismatch if someone had copied a section of some other image and pasted it on the top of this one to hide what is underneath.

( IV ) When we apply Error Level Analysis (ELA) [135] to Fig. 23a, we see significant differences in the processed image in brightness levels again in the same places, Fig. 28. For comparison, we also provide the analysis of Fig. 23b; the output is in Fig. 28b.



**Figure 28:** ELA of Fig. 23a (a) and Fig. 23b (b) [41].

In conclusion, there is no doubt whatsoever that the 23a image is merely the original image with these changes made:

- The overfilled part of the root canal was added by drawing with a grey color.
- The rest of the root canals have their colors smoothed out, to blend in with the newly added area.
- The white band in the lower right corner of the original image was removed by copying parts of a tooth from another image; the cut corners were similarly drawn over.

As seen from Section 7, our department has close cooperation with the Police of the Czech Republic. Based on this example, we are preparing a new project for *Forensic dentistry*: preparing an automatic tool for matching X-ray images from a doctor's database to an X-ray image supplied by the police. In this way we hope to help with identification of found human remains.



## Part II

### MAIN PART OF THE THESIS



GOALS

---

## IMAGE FORENSICS:

- Detection of image re-capturing
- Copy-move forgery detection in tampered images
- Construction of a high-performance camera sensor fingerprint
- Design a framework for forensic image and video analysis

## MEDICAL IMAGING:

- Detection of region of interest in Wireless Capsule Endoscopy ([WCE](#)) images
- Development of preprocessing and enhancing methods for Videokymography ([VKG](#)) data
- Design and validation of framework for complex vocal evaluation



## STRUCTURE OF THE THESIS

---

The Thesis consists of six papers that are attached below. Each of them presents the work that has contributed to achieving the main goals. The first four papers focus on *Image Forensics*, a set of methods and developing new approaches. The last two papers focus on medical image data, their analysis, segmentation and calculation of features.

The first paper consists of an overview of the developed and implemented tools for image forensic analysis at our department, which were used in PIZZARO software. The PIZZARO originated in collaboration with the Institute of Criminalistics, National Drug Headquarters of the Criminal Police and Investigation Service of the Police of the Czech Republic. It was designed to help to assess image and video credibility and origin and to restore and increase image quality by diminishing unwanted blur, noise, and other possible artifacts.

In the second paper we address the problem of identification of pictures and videos re-captured from LCD screens. We show that they often exhibit detectable periodic patterns that are caused by the regular sampling grid of LCD screen and aliasing. We develop a novel method capable of detecting these patterns by using the theory of cyclostationarity. The results show a higher success rate than people.

In the third conference paper we focus on the question how many reference images should be used to construct a high-performance camera sensor fingerprint. We find that about 50 and more images are recommended in the literature, the same as in the original article introducing this issue. However, we get and present other results from our experiments.

In the fourth paper we attempt to combat copy-move forgery in JPEG images, where the compression significantly degrades detection by standard algorithms. We derive a JPEG-based constraint that any pair of patches must satisfy to be considered a valid candidate for a tampering area. We also propose an efficient algorithm to verify the constraint that can be integrated into most of the existing methods. Experiments show significant improvement of detection, especially for difficult cases, such as small objects, objects covered by textureless areas and repeated patterns, where it reduces false detec-

tion greatly.

The fifth paper describes a new way of displaying bleeding spots in the small intestine. We propose two automatic methods for detecting bleeding in [WCE](#) videos. The first one uses solely color information, whereas the second one incorporates assumptions about the blood spot shape and size. Both methods can be applied either individually or their results can be fused together for the final decision.

The last conference paper proposes methods for computer-assisted evaluation of diagnostically important vibration features, related to movements of vocal folds and their surroundings. They are derived from existing as well as newly developed methods of digital *Image Processing*, mainly based on data segmentation and morphological operations. Performance of the developed methods is compared to expert manual assessments and it proves to be comparable with clinicians' conclusions.

Brief summaries of the attached papers are in the following sections in the same order.

## PIZZARO: FORENSIC ANALYSIS AND RESTORATION OF IMAGE AND VIDEO DATA

---

### CITATION

J. Kamenický et al. "PIZZARO: Forensic analysis and restoration of image and video data." In: *Forensic Science International* 264 (2016). Special Issue on the 7th European Academy of Forensic Science Conference, pp. 153 –166. ISSN: 0379-0738. DOI: [10.1016/j.forsciint.2016.04.027](https://doi.org/10.1016/j.forsciint.2016.04.027)

Shorter version also appeared at:

B. Zitová et al. "Imaging device identification and detection of image tampering." In: *Abstract book of 7th European Academy of Forensic Science Conference*. Prague, Czechia, Sept. 2015

### ABSTRACT

This paper introduces a set of methods for image and video forensic analysis. They were designed to help to assess image and video credibility and origin and to restore and increase image quality by diminishing unwanted blur, noise, and other possible artifacts. The motivation came from the best practices used in the criminal investigation utilizing images and/or videos. The determination of the image source, the verification of the image content, and image restoration were identified as the most important issues of which automation can facilitate criminalists work. Novel theoretical results complemented with existing approaches (LCD re-capture detection and denoising) were implemented in the PIZZARO software tool, which consists of the image processing functionality as well as of reporting and archiving functions to ensure the repeatability of image analysis procedures and thus fulfills formal aspects of the image/video analysis work. Comparison of new proposed methods with the state of the art approaches is shown. Real use cases are presented, which illustrate the functionality of the developed methods and demonstrate their applicability in different situations. The use cases as well as the method design were solved in tight cooperation of scientists from the Institute of Criminalistics, National Drug Headquarters of the Criminal Police and Investigation Service of the Police of the Czech Republic, and image processing experts from the Czech Academy of Sciences.

## MAIN CONTRIBUTION OF THE PAPER



**Figure 29:**  
Three modules of  
PIZZARO.

In this paper, we introduce a set of methods for image and video forensic analysis. We address two key problems identified as the most important areas in which automation can facilitate criminalists' work. They are verification of image/video data with respect to their credibility and origin, and image restoration aimed at diminishing unwanted artifacts and at the same time increasing the data quality using all available information.

The approaches are described for source device identification, the image content verification methods based on various image characteristics, and finally, methodology for removing compression artifacts and even increasing the data resolution.

The functionality of some methods is illustrated by comparison to other existing approaches and by demonstrating performance tests. Additionally, the detection of **LCD** image re-capture and noise removal are implemented in order to provide required functionality.

Proposed methods were designed or modified in order to fulfill time, quality, and interactivity criteria. It should be noted that the output of the methods often has an indicative nature, thus the recommended practice is to apply several methods to get stronger evidence.

All theoretical results were implemented in the PIZZARO software tool (available for testing and licensing), which consists of the *Image Processing* functionality and reporting and archiving functions to ensure the repeatability of image analysis procedures, fulfilling formal aspects of the analysis work. This tool is used by the Police of the Czech Republic, the Institute of Criminalistics, the National Drug Headquarters of the Criminal Police and the Investigation Service of the Police of the Czech Republic. At the international level we have users with registration from China, Germany, Georgia, India, Indonesia, and Spain.

## MAIN CONTRIBUTION OF THE AUTHOR

- New methods for *Image source determination* and *Image content verification* modules:
  - Assignment to the source device
  - **LCD** image re-capture
  - Detection of copy-move
- Experiments: design, data preparation, realization, and interpretation
- PIZZARO **GUI** and web support design and implementation

## DETECTING CYCLOSTATIONARITY IN RE-CAPTURED LCD SCREENS

### CITATION

B. Mahdian, A. Novozámský, and S. Saic. “Detecting Cyclostationarity in Re-Captured LCD Screens.” In: *Journal of Forensic Research* 6.4 (2015), pp. 1–6. ISSN: 2157-7145. DOI: [10.4172/2157-7145.1000294](https://doi.org/10.4172/2157-7145.1000294)

Shorter version also appeared at:

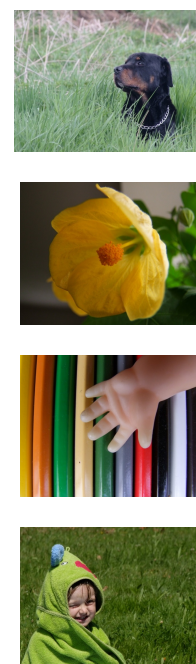
B. Mahdian, A. Novozámský, and S. Saic. “Identification of aliasing-based patterns in re-captured LCD screens.” In: *2015 IEEE International Conference on Image Processing (ICIP)*. Québec City, Canada, Sept. 2015, pp. 616–620. DOI: [10.1109/ICIP.2015.7350872](https://doi.org/10.1109/ICIP.2015.7350872)

### ABSTRACT

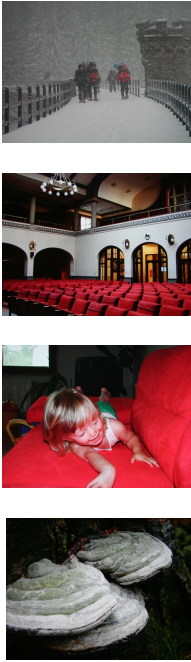
It is easy to display a video or picture on an LCD screen and recapture it by using a camera to hide traces of digital image manipulation or fool an access system based on face recognition technique. In this paper, we show that humans do not have a good performance in detecting recaptured data from LCD screens. Hence, it is important to have methods capable of distinguishing between natural videos and pictures and those recaptured ones. In this paper, we show that, typically, recaptured images and videos from LCD screens exhibit detectable periodic patterns that are caused by regular sampling grid of the LCD monitor and aliasing. We develop our method using the theory of cyclostationarity and experimentally validate it. The term cyclostationarity refers to a special class of signals which exhibit periodicity in their statistics. Our method will be based on the fact that a cyclostationary signal has a frequency spectrum correlated with a shifted version of itself.

### MAIN CONTRIBUTION OF THE PAPER

In this paper we address the problem of identification of pictures and videos re-captured from LCD screens. The motivation for detecting re-captured images and videos can come from several sources. One practical use might be the automatic detection of illegally captured movies from LCD screens. Re-capturing is an easy tool to eliminate copyright-related invisible watermarks hidden in images and videos.



**Figure 30:**  
Examples of  
natural pictures  
from our  
experiment.



**Figure 31:**  
Examples of  
re-captured  
pictures from our  
experiment.

Another area that needs to be capable of detecting re-captured videos and images is the authentication area. Access systems using face recognition techniques are often vulnerable to spoofing attacks. In a spoofing attempt, a person tries to masquerade as another person to gain access to the system. Our motivation comes from the digital forensics point of view. We show that re-captured pictures often exhibit detectable periodic patterns that are caused by the regular sampling grid of LCD screens and aliasing. We develop a method capable of detecting these patterns by using the theory of cyclostationarity. The term cyclostationarity refers to a special class of signals which exhibit periodicity in their statistics. Such signals have a frequency spectrum correlated with a shifted version of itself. Experimental results quantifying the performance of the developed method are also shown.

A big part of the paper is a survey-based experiment in how good humans are at identifying re-captured images from an LCD screen. For this purpose we created a web application where respondents rated pictures as LCD re-captured or natural pictures, from a total of 800 images. The results confirmed that humans do not have a good performance in detecting re-captured data from LCD screens. Hence, it is important to have methods capable of distinguishing between natural videos and pictures and those re-captured ones.

#### MAIN CONTRIBUTION OF THE AUTHOR

- Collaboration on the development of the proposed method
- Creation of the ground-true data set: web based application for the online test, how good are humans at identifying LCD re-captured images; statistical evaluation
- Experiments: design, data preparation, realization, and interpretation

## DETERMINATION OF STOP-CRITERION FOR CONSTRUCTING SENSOR FINGERPRINT

### CITATION

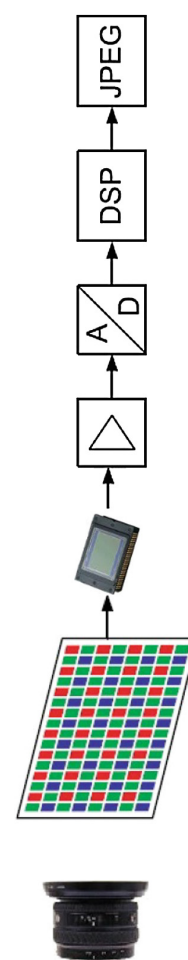
B. Mahdian, A. Novozámský, and S. Saic. "Determination of Stop-Criterion for Incremental Methods Constructing Camera Sensor Fingerprint." In: *Digital-Forensics and Watermarking: 13th International Workshop, IWDW 2014*. Ed. by Y.-Q. Shi et al. Taipei, Taiwan: Springer International Publishing, Oct. 2015, pp. 47–59. ISBN: 978-3-319-19321-2. DOI: [10.1007/978-3-319-19321-2\\_4](https://doi.org/10.1007/978-3-319-19321-2_4)

### ABSTRACT

This paper aims to find the minimum sample size of the camera reference image set that is needed to build a sensor fingerprint of a high performance. Today's methods for building sensor fingerprints do rely on having a sufficient number of camera reference images. But, there is no clear answer to the question of how many camera reference images are really needed? In this paper, we will analyze and find out how to determine the minimum needed number of reference images to remove the mentioned uncertainty. We will introduce a quantitative measure (a stop-criterion) stating how many photos should be used to create a high-performance sensor fingerprint. This stop-criterion will directly reflect the confidence level that we would like to achieve. By considering that the number of digital images used to construct the camera sensor fingerprint can have a direct impact on performance of the sensor fingerprint, it is apparent that this, so far underestimated, topic is of major importance.

### MAIN CONTRIBUTION OF THE PAPER

To decide whether a digital image or video has been captured by a particular imaging device, a sensor fingerprint of the device is typically constructed. This is usually achieved by averaging sensor noise obtained from a set of reference images captured by the device under investigation [44]. But how many reference images should be used to construct a well-performing and reliable sensor fingerprint? Should we use 10, 50, or 999 images? The cardinality of the camera reference set has an essential role in the performance of the constructed sensor fingerprint. An insufficient number of reference images leads to failing to achieve a successful image (video) ballistics test.



**Figure 32:**  
A typical digital  
camera system.

It is clear that this question is of major importance and there is a need to remove the uncertainty around it. In this paper, we show that the cardinality must be determined by using a stop-criterion or metric adopted and customized to the particular camera and reference set. To this end, we introduce a stop-criterion measuring the convergence of sensor additive noise to zero with respect to the growth of cardinality of the camera reference set. To achieve our goals, we also analytically show a novel way of estimating the systematic noise component of the imaging sensor that forms the sensor fingerprint. The theoretical model is built by using the multiplicative nature of the noise component.

#### MAIN CONTRIBUTION OF THE AUTHOR

- Collaboration on the development and implementation of the proposed method
- Extensive dataset design and preparation
- Experiments: design, realization, and interpretation

## COPY-MOVE FORGERY DETECTION USING JPEG COMPRESSION MODEL

### CITATION

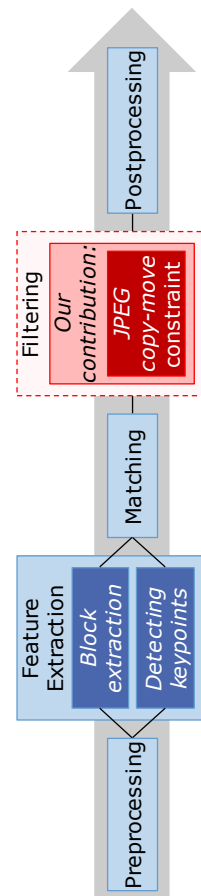
A. Novozámský and M. Šorel. "Copy-move forgery detection using JPEG compression model." In: *Submitted to Forensic Science International* (2017)

### ABSTRACT

The copy-move forgery, based on copying an object and pasting in another location of the same image, is a common way to manipulate image content. In this paper, we address the problem of copy-move forgery detection in JPEG images. The main problem with JPEG compression is that the same pixels, after moving to a different position and storing in the JPEG format, have different values. The majority of existing algorithms is based on matching pairs of similar patches, which generates many false matches. In many cases they cannot be eliminated by postprocessing, causing the failure of detection. To overcome this problem, we derive a JPEG-based constraint that any pair of patches must satisfy to be considered a valid candidate and propose an efficient algorithm to verify the constraint. The constraint can be integrated into most existing methods. Experiments show significant improvement of detection, especially for difficult cases, such as small objects, object covered by textureless areas and repeated patterns.

### MAIN CONTRIBUTION OF THE PAPER

In this paper, we propose a method to improve the reliability of detecting copy-move forgery in JPEG images. For this purpose, we introduce the JPEG copy-move constraint that can be used to filter out false matches of candidate patches in almost any existing detection algorithm. Since the constraint is exact, i.e. gives almost no false negatives, results are always better than without this constraint. The typical processing pipeline of copy-move forgery detection algorithms is demonstrated in a diagram on the right side of this page. The main contribution of this paper is filtering using the JPEG copy-move constraint, which increases sensitivity and significantly reduces the number of false matches.



**Figure 33:**  
Typical processing  
pipeline of  
copy-move forgery  
detection  
algorithms.

To demonstrate how our procedure improves performance of a complete detection algorithm, we use simple patch-level matching, followed by the [JPEG](#) copy-move constraint and verification of the coherence of shift vectors. To simplify the comparison to other methods mentioned in literature, we use two standard image datasets [[123](#), [133](#)] and two methods based on different principles. We also created a third image set containing difficult cases with repeated patterns, objects masked by an indistinct texture (skies, building facades, etc.) and small objects. The performance of other methods on the same datasets is analyzed in [[21](#), [123](#)].

To evaluate the algorithms, we use two different metrics. In both cases, we compare the output of algorithms with the ground truth represented, for each image, by a matrix of ones on the tampered pixels and zeroes elsewhere. The results of experiments shows that our proposed method improves results significantly even when used with the simple patch-level matching distance between patches calculated by the  $l_2$  norm.

The proposed approach cannot be directly used with geometrical transforms like scale change or rotation. One could imagine an extension to these operations but the results probably would not be worth the additional complexity of the algorithm. In our opinion, improved detection justifies this restriction, though. In addition, the scenario of pure shift is probably frequent enough to be considered separately. Another limitation of the [JPEG](#) constraint is that it requires the moved object to contain at least one [JPEG](#) block. On the other hand, smaller objects are probably too hard also for all other methods.

#### MAIN CONTRIBUTION OF THE AUTHOR

- Development of the new method
- Creation of the ground-true data set
- Experiments: design, data preparation, realization, and interpretation

## AUTOMATIC BLOOD DETECTION IN CAPSULE ENDOSCOPY VIDEO

### CITATION

A. Novozámský et al. "Automatic blood detection in capsule endoscopy video." In: *Journal of Biomedical Optics* 21.12 (2016), p. 126007. DOI: [10.1117/1.JBO.21.12.126007](https://doi.org/10.1117/1.JBO.21.12.126007)

### ABSTRACT

We propose two automatic methods for detecting bleeding in wireless capsule endoscopy videos of the small intestine. The first one uses solely the color information, whereas the second one incorporates the assumptions about the blood spot shape and size. The original idea is namely the definition of a new color space that provides good separability of blood pixels and intestinal wall. Both methods can be applied either individually or their results can be fused together for the final decision. We evaluate their individual performance and various fusion rules on real data, manually annotated by an endoscopist.

### MAIN CONTRIBUTION OF THE PAPER

**WCE** is a noninvasive diagnostic tool for the small bowel. The patient swallows a cylindrical plastic capsule measuring 10 x 25 mm containing a camera, **LED** light source, signal transmitter and battery. The main obstacle to routine usage of **WCE** is that the visual evaluation of the video is very time consuming. One examination produces up to 60,000 images, which a trained endoscopist has to check. The main contribution of our paper is a semi-automatic technique which significantly reduces the evaluation time while the final decision is still left to the endoscopist.

We propose two automatic methods for detecting bleeding in **WCE** video. The first one uses solely the color information. The original idea is namely the definition of a new color space that provides good separability of blood pixels and the intestinal wall. The second one incorporates the assumptions about the blood spot shape and size. We tested both methods individually as well as in various combinations. We evaluated the results on a large test set, manually annotated by an endoscopist.

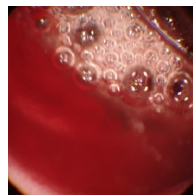
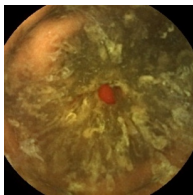
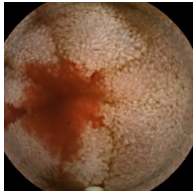
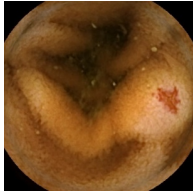


**Figure 34:**  
Examples of  
frames without  
blood.

We believe that the presented technique significantly saves the time of endoscopists required for visual expert assessment. The method is currently in use at the University Hospital of Charles University in Hradec Králové, Czech Republic. We would like to extend the method to detect ulcers and tumors in the future.

#### MAIN CONTRIBUTION OF THE AUTHOR

- Proposal and implementation of the novel method
- Analysis and testing of existing methods and softwares
- Design and implementation of video handling procedures
- Creation of ground truth dataset (ill/healthy data)
- Experiments: design, data preparation, realization, and interpretation
- Design of the user-friendly GUI for doctors



**Figure 35:**  
*Examples of  
frames with blood.*

## IMAGE ANALYSIS OF VIDEOKYMOGRAPHIC DATA

## CITATION

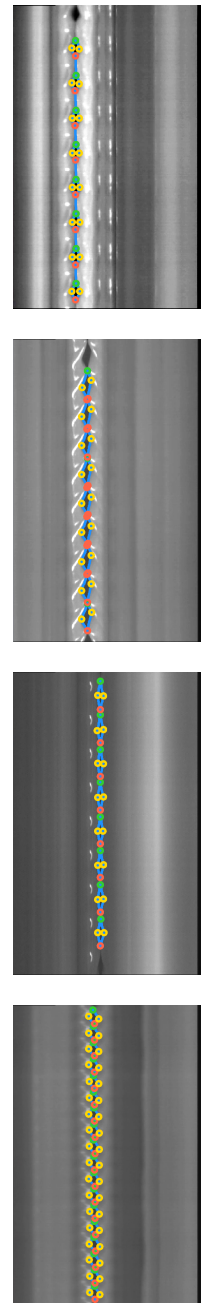
A. Novozámský et al. "Image analysis of videokymographic data." In: *2015 IEEE International Conference on Image Processing (ICIP)*. Québec City, Canada, Sept. 2015, pp. 78–82. DOI: [10.1109/ICIP.2015.7350763](https://doi.org/10.1109/ICIP.2015.7350763)

## ABSTRACT

Videokymography (VKG) is a high-speed medical imaging technique used in laryngology and phoniatrics for examination of vocal fold vibrations, it offers important characteristics for diagnosis and treatment of voice disorders. VKG repeatedly scans only a single line from the scene and captures movements of vocal folds in this region of interest. This paper proposes methods for computer assisted evaluation of diagnostically important vibration features, related to movements of vocal folds and their surroundings. They are derived from existing as well as newly developed methods of digital image processing, mainly based on data segmentation and morphological operations. Performance of the developed methods is compared to expert manual assessments and it proves to be comparable with clinicians conclusions.

## MAIN CONTRIBUTION OF THE PAPER

Our paper proposes a software solution for the enhancement of videokymographic data and the extraction of typical characteristics of vocal folds vibrations, such as the time-varying extent of rima glottidis and the progression of mucosal waves, and corresponding vibration parameters (e.g., frequency, symmetries, amplitude). The proposed methodology is based on well-established digital *Image Processing* methods; namely, image denoising, edge detection, image segmentation, and object identification. The set of evaluated features was designed by experts, drawing on their longtime expertise with videokymography. These features are mostly based on the estimated shape of the glottal contour, i.e., the boundary between the vocal folds and the rima glottidis, but also on mucosal waves. We trace a set of base features—opening and closing points and ventral and medial peaks. They determine the duration of the vibration cycle and the symmetries. They are used for deriving the shape, the amplitudes, and the



**Figure 36:**  
Analyzed frames  
of VKG.

extent of mucosal waves, too. Accurate detection and measurement of these base features is critical for proper computation of further vocal fold vibration parameters. In addition to the automated detection of analysis parameters, the proposed software offers clinicians an option to manually intervene in any algorithm step.

A representative set of videokymographic data with a wide range of vibration patterns was used for testing the developed system (50 videokymograms evaluated by 18 expert and non-expert raters). The performance of the new software was assessed by comparing the outcome of the automated parameter detection to the manual/visual ratings. The proposed software tool, Videokymography Feature Detection Software ([VKFD](#)), was developed for computerized analysis of videokymographic data, but also allows direct inputs from clinicians. Such computer-assisted feature extraction and parameter estimation will facilitate the diagnostic process. Moreover, it will produce parameter sets, which can be numerically compared to previous analyses.

#### MAIN CONTRIBUTION OF THE AUTHOR

- Proposal and implementation of the novel methods for:
  - selection of information-rich video segments
  - detection of vocal fold borderline
  - features of interest computation
- Analysis of existing methods
- Experiments: design, data preparation, realization, and interpretation
- Cooperation with Austrian colleague on the C++ framework implementation
- Design of the user-friendly [GUI](#) for doctors

## MAIN CONTRIBUTION OF THE THESIS

---

This Thesis contributes to two areas of *Image Processing*. Specifically, it deals with methods and techniques for forensics science and medical imaging. The main text consists of six papers describing novel approaches and their major contributions are summarized below:

- Two methods for *Image Forensics* were developed, tested and reported. The new approach was applied to the copy-move forgery detection and the technique for finding out the [LCD](#) re-capturing was described.
- The efficient methodology was introduced to extract the camera sensor fingerprint.
- The complex software tool was created to assess authenticity of image data for forensic purposes.
- A new technique for fast blood detection in Wireless Capsule Endoscopy ([WCE](#)) was proposed and incorporated into the original user-friendly software for gastroenterologists.
- New methods for preprocessing the Videokymography ([VKG](#)) data, detection of vocal borderline, and selection of information-rich video segments were proposed. The C++ framework for analyzing the [VKG](#) data and unique [GUI](#) for phoniatrists was developed.



## BIBLIOGRAPHY

---

- [1] *A brief history of CT*. <http://www.impactscan.org/CThistory.htm>. (Accessed on 08/08/2017).
- [2] *A brief history of forensic radiology — University of Leicester*. <http://www2.le.ac.uk/departments/emfpu/imaging/brief-history>. (Accessed on 08/25/2017).
- [3] *"A picture is worth a thousand words" - the meaning and origin of this phrase*. <http://www.phrases.org.uk/meanings/a-picture-is-worth-a-thousand-words.html>. (Accessed on 08/25/2017).
- [4] I. Amerini, R. Caldelli, V. Cappellini, F. Picchioni, and A. Piva. "Analysis of denoising filters for photo response non uniformity noise extraction in source camera identification." In: *16th International Conference on Digital Signal Processing*. 2009, pp. 1–7. DOI: [10.1109/ICDSP.2009.5201240](https://doi.org/10.1109/ICDSP.2009.5201240).
- [5] T. Baar, W. van Houten, and Z. J. M. H. Geradts. "Camera identification by grouping images from database, based on shared noise patterns." In: *CoRR* abs/1207.2641 (2012). URL: <http://arxiv.org/abs/1207.2641>.
- [6] S. Bayram, H. T. Sencar, and N. Memon. "Efficient techniques for sensor fingerprint matching in large image and video databases." In: *Proc. SPIE 7541, Media Forensics and Security II*. Vol. 7541. 2010, pp. 754109–754109–8. DOI: [10.1117/12.845737](https://doi.org/10.1117/12.845737).
- [7] S. Bayram, H. T. Sencar, and N. Memon. "Efficient Sensor Fingerprint Matching Through Fingerprint Binarization." In: *IEEE Transactions on Information Forensics and Security* 7.4 (2012), pp. 1404–1413. ISSN: 1556-6013. DOI: [10.1109/TIFS.2012.2192272](https://doi.org/10.1109/TIFS.2012.2192272).
- [8] P. Bestagini, M. Visentini-Scarzanella, M. Tagliasacchi, P. L. Dragotti, and S. Tubaro. "Video recapture detection based on ghosting artifact analysis." In: *2013 IEEE International Conference on Image Processing (ICIP)*. 2013, pp. 4457–4461. DOI: [10.1109/ICIP.2013.6738918](https://doi.org/10.1109/ICIP.2013.6738918).
- [9] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh. "Computationally Efficient Face Spoofing Detection with Motion Magnification." In: *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2013, pp. 105–110. DOI: [10.1109/CVPRW.2013.23](https://doi.org/10.1109/CVPRW.2013.23).

- [10] I. Bjarnason, J. Hayllar, A. J. Macpherson, and A. S. Russell. "Side effects of nonsteroidal anti-inflammatory drugs on the small and large intestine in humans." In: *Gastroenterology* 104.6 (1993), pp. 1832–1847. ISSN: 0016-5085. DOI: [http://dx.doi.org/10.1016/0016-5085\(93\)90667-2](http://dx.doi.org/10.1016/0016-5085(93)90667-2).
- [11] G. J. Bloy. "Blind Camera Fingerprinting and Image Clustering." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.3 (2008), pp. 532–534. ISSN: 0162-8828. DOI: [10.1109/TPAMI.2007.1183](https://doi.org/10.1109/TPAMI.2007.1183).
- [12] X. Bo, W. Junwen, L. Guangjie, and D. Yuewei. "Image Copy-Move Forgery Detection Based on SURF." In: *2010 International Conference on Multimedia Information Networking and Security*. 2010, pp. 889–892. DOI: [10.1109/MINES.2010.189](https://doi.org/10.1109/MINES.2010.189).
- [13] W. C. Bradford. *Reaching the visual learner: teaching property through art*. The Law Teacher Vol. 11, 2004. <https://ssrn.com/abstract=587201>. (Accessed on 08/25/2017).
- [14] H. Cao and A. C. Kot. "Identification of recaptured photographs on LCD screens." In: *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2010, pp. 1790–1793. DOI: [10.1109/ICASSP.2010.5495419](https://doi.org/10.1109/ICASSP.2010.5495419).
- [15] G. Cattaneo, P. Faruolo, and U. F. Petrillo. "Experiments on Improving Sensor Pattern Noise Extraction for Source Camera Identification." In: *2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*. 2012, pp. 609–616. DOI: [10.1109/IMIS.2012.35](https://doi.org/10.1109/IMIS.2012.35).
- [16] G. Cattaneo, G. Roscigno, and U. F. Petrillo. "A Scalable Approach to Source Camera Identification over Hadoop." In: *2014 IEEE 28th International Conference on Advanced Information Networking and Applications*. 2014, pp. 366–373. DOI: [10.1109/AINA.2014.47](https://doi.org/10.1109/AINA.2014.47).
- [17] L.-H. Chan, N.-F. Law, and W.-Ch. Siu. "A two dimensional camera identification method based on image sensor noise." In: *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. 2012, pp. 1741–1744. DOI: [10.1109/ICASSP.2012.6288235](https://doi.org/10.1109/ICASSP.2012.6288235).
- [18] C. Chen, J. Ni, Z. Shen, and Y. Q. Shi. "Blind Forensics of Successive Geometric Transformations in Digital Images Using Spectral Method: Theory and Applications." In: *IEEE Transactions on Image Processing* 26.6 (2017), pp. 2811–2824. ISSN: 1057-7149. DOI: [10.1109/TIP.2017.2682963](https://doi.org/10.1109/TIP.2017.2682963).
- [19] M. Chen, J. Fridrich, M. Goljan, and J. Lukáš. "Source digital camcorder identification using sensor photo response non-uniformity." In: *Proceeding of the SPIE, 2007*. Vol. 6505. 1. SPIE, 2007.

- [20] W. Chen, Y. Q. Shi, and W. Su. "Image splicing detection using 2-D phase congruency and statistical moments of characteristic function." In: vol. 6505. 2007, 65050R–65050R–8. DOI: [10.1117/12.704321](https://doi.org/10.1117/12.704321). URL: <http://dx.doi.org/10.1117/12.704321>.
- [21] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou. "An Evaluation of Popular Copy-Move Forgery Detection Approaches." In: *IEEE Transactions on Information Forensics and Security* 7.6 (2012), pp. 1841–1854. ISSN: 1556-6013. DOI: [10.1109/TIFS.2012.2218597](https://doi.org/10.1109/TIFS.2012.2218597).
- [22] H. T. Chugani, M. E. Phelps, and J. C. Mazziotta. "Positron emission tomography study of human brain functional development." In: *Annals of Neurology* 22.4 (1987), pp. 487–497. ISSN: 1531-8249. DOI: [10.1002/ana.410220408](https://doi.org/10.1002/ana.410220408). URL: <http://dx.doi.org/10.1002/ana.410220408>.
- [23] V. Conotter and G. Boato. "Analysis of sensor fingerprint for source camera identification." In: *Electronics letters* 47.25 (Jan. 1, 2011), pp. 1366–1367.
- [24] A. J. Cooper. "Improved photo response non-uniformity (PRNU) based source camera identification." In: *Forensic Science International* 226.1 (2013), pp. 132–141. ISSN: 0379-0738. DOI: <http://dx.doi.org/10.1016/j.forsciint.2012.12.018>.
- [25] J. R. Corripio, A. L. S. Orozco, L. J. G. Villalba, J. Hernandez-Castro, and S. J. Gibson. "Source Smartphone Identification Using Sensor Pattern Noise and Wavelet Transform." unpublished. 2013. URL: <http://kar.kent.ac.uk/37195/>.
- [26] A. Cortiana, V. Conotter, G. Boato, and F. G. B. De Natale. "Performance comparison of denoising filters for source camera identification." In: *Proc. SPIE*. Vol. 7880. 2011, pp. 788007–788007–6. DOI: [10.1117/12.872489](https://doi.org/10.1117/12.872489).
- [27] *Culture-DAT BASE*. <https://database.wordpress.com/category/culture/page/2/>. (Accessed on 08/15/2017).
- [28] S. Dehnie, T. Sencar, and N. Memon. "Digital Image Forensics for Identifying Computer Generated and Digital Camera Images." In: *2006 International Conference on Image Processing (ICIP)*. 2006, pp. 2313–2316. DOI: [10.1109/ICIP.2006.312849](https://doi.org/10.1109/ICIP.2006.312849).
- [29] *Detecting Forged Images*. <http://www.smtdp.com/detecting-forged-images/>. (Accessed on 08/15/2017).
- [30] A. E. Dirik and A. Karakucuk. "Forensic use of photo response non-uniformity of imaging sensors and a counter method." In: *Opt Express* 22.1 (2014), pp. 470–482.

- [31] L. Du and H. Ling. "Cross-age face verification by coordinating with cross-face age verification." In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 2329–2338. DOI: [10.1109/CVPR.2015.7298846](https://doi.org/10.1109/CVPR.2015.7298846).
- [32] K. T. Dussik. "Weitere Ergebnisse der Ultraschalluntersuchung bei Gehirnerkrankungen." In: *Acta Neurochirurgica* 2.3 (1952), pp. 379–401. ISSN: 0942-0940. DOI: [10.1007/BF01405828](https://doi.org/10.1007/BF01405828). URL: <https://doi.org/10.1007/BF01405828>.
- [33] R. Eliakim et al. "A Prospective Study of the Diagnostic Accuracy of PillCam ESO Esophageal Capsule Endoscopy Versus Conventional Upper Endoscopy in Patients With Chronic Gastroesophageal Reflux Diseases." In: *Journal of Clinical Gastroenterology* 39.7 (2005). ISSN: 0192-0790.
- [34] *Endodontics Dentists*. <http://www.currydentistry.com/endodontics.php>. (Accessed on 08/11/2017).
- [35] J. Fan, H. Cao, and A. C. Kot. "Estimating EXIF Parameters Based on Noise Features for Image Manipulation Detection." In: *IEEE Transactions on Information Forensics and Security* 8.4 (2013), pp. 608–618. ISSN: 1556-6013. DOI: [10.1109/TIFS.2013.2249064](https://doi.org/10.1109/TIFS.2013.2249064).
- [36] H. Farid. "Image forgery detection." In: *IEEE Signal Processing Magazine* 26.2 (2009), pp. 16–25. ISSN: 1053-5888. DOI: [10.1109/MSP.2008.931079](https://doi.org/10.1109/MSP.2008.931079).
- [37] H. Farid. "Image forgery detection." In: *IEEE Signal Processing Magazine* 26.2 (2009), pp. 16–25. ISSN: 1053-5888. DOI: [10.1109/MSP.2008.931079](https://doi.org/10.1109/MSP.2008.931079).
- [38] P. Ferrara, T. Bianchi, A. De Rosa, and A. Piva. "Image Forgery Localization via Fine-Grained Analysis of CFA Artifacts." In: *IEEE Transactions on Information Forensics and Security* 7.5 (2012), pp. 1566–1577. ISSN: 1556-6013. DOI: [10.1109/TIFS.2012.2202227](https://doi.org/10.1109/TIFS.2012.2202227).
- [39] R. A. Fessenden. *Signaling by sound and other longitudinal elastic impulses*. US Patent 1,108,895. 1914. URL: <https://www.google.com/patents/US1108895>.
- [40] *First photograph, View from the Window at Le Gras, Joseph Nicéphore Niépce, ca. 1826*. [http://www.hrc.utexas.edu/exhibitions/permanent/windows/southeast/joseph\\_nicephore\\_niepce.html](http://www.hrc.utexas.edu/exhibitions/permanent/windows/southeast/joseph_nicephore_niepce.html). (Accessed on 08/28/2017).
- [41] *FotoForensics*. <http://fotoforensics.com/tutorial-ela.php>. (Accessed on 08/08/2017).
- [42] J. Fridrich. "Digital image forensics." In: *IEEE Signal Processing Magazine* 26.2 (2009), pp. 26–37. ISSN: 1053-5888. DOI: [10.1109/MSP.2008.931078](https://doi.org/10.1109/MSP.2008.931078).

- [43] J. Fridrich, D. Soukal, and J. Lukáš. "Detection of Copy Move Forgery in Digital Images." In: *Digital Forensic Research Workshop*. Cleveland, OH, Aug. 2003.
- [44] J. Fridrich, D. Soukal, and J. Lukáš. "Detection of copy-move forgery in digital images." In: *Proceedings of Digital Forensic Research Workshop*. Cleveland, OH, USA: IEEE Computer Society, 2003, pp. 55–61.
- [45] A. C. Gallagher. "Detection of linear and cubic interpolation in JPEG compressed images." In: *The 2nd Canadian Conference on Computer and Robot Vision*. 2005, pp. 65–72. DOI: [10.1109/CRV.2005.33](https://doi.org/10.1109/CRV.2005.33).
- [46] W. A. Gardner, A. Napolitano, and L. Paura. "Cyclostationarity: Half a century of research." In: *Signal Processing* 86.4 (2006), pp. 639–697. ISSN: 0165-1684. DOI: <http://dx.doi.org/10.1016/j.sigpro.2005.06.016>.
- [47] *Given Imaging*. <http://www.givenimaging.com/Pages/default.html>. (Accessed on 08/09/2017).
- [48] H. Gohe and T. Wedekind. "Der Ultraschall in der Medizin." In: *Klinische Wochenschrift* 19.2 (1940), pp. 25–29. ISSN: 1432-1440. DOI: [10.1007/BF01772492](https://doi.org/10.1007/BF01772492). URL: <https://doi.org/10.1007/BF01772492>.
- [49] M. Goljan and J. Fridrich. "Camera identification from cropped and scaled images." In: *Proc. SPIE, Electronic Imaging, Forensics, Security, Steganography, and Watermarking of Multimedia Contents X*. Vol. 6819. 2008, 68190E–1–68190E–13. DOI: [10.1117/12.766732](https://doi.org/10.1117/12.766732).
- [50] M. Goljan and J. Fridrich. "Sensor fingerprint digests for fast camera identification from geometrically distorted images." In: vol. 8665. 2013, 86650B–1–86650B–10. DOI: [10.1117/12.2003234](https://doi.org/10.1117/12.2003234).
- [51] M. Goljan, J. Fridrich, and T. Filler. "Large Scale Test of Sensor Fingerprint Camera Identification." In: *SPIE Conference on Media Forensics and Security*. 2009.
- [52] J. Hartloff, M. Morse, B. Zhang, T. Effland, J. Cordaro, J. Schuler, S. Tulyakov, A. Rudra, and V. Govindaraju. "A multiple server scheme for fingerprint fuzzy vaults." In: *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2015, pp. 119–127. DOI: [10.1109/CVPRW.2015.7301327](https://doi.org/10.1109/CVPRW.2015.7301327).
- [53] W. Herriott and D. W. Farnsworth. "High Speed Motion Pictures of the Vocal Cords." In: *The Journal of the Acoustical Society of America* 9.3 (1938), pp. 274–274. DOI: [10.1121/1.1902052](https://doi.org/10.1121/1.1902052).
- [54] M. Hirano. *Clinical examination of voice*. Disorders of human communication. Springer London, 1981. ISBN: 9783211816592.

- [55] *History of Ultrasound in Obstetrics and Gynecology, Part 1*. <http://www.ob-ultrasound.net/history1.html>. (Accessed on 08/09/2017).
- [56] Y. Hu, C.-T. Li, and C. Jian. "Building Fingerprints with Information from Three Color Bands for Source Camera Identification." In: *Proceedings of the 2Nd ACM Workshop on Multimedia in Forensics, Security and Intelligence*. MiFor '10. Firenze, Italy: ACM, 2010, pp. 111–116. ISBN: 978-1-4503-0157-2. DOI: [10.1145/1877972.1878001](https://doi.org/10.1145/1877972.1878001).
- [57] Y. Hu, C.-T. Li, and Z. Lai. "Fast source camera identification using matching signs between query and reference fingerprints." In: *Multimedia Tools and Applications* 74.18 (2015), pp. 7405–7428. ISSN: 1573-7721. DOI: [10.1007/s11042-014-1985-3](https://doi.org/10.1007/s11042-014-1985-3).
- [58] G. J. Iddan, D. Avni, A. Glukhovsky, and G. Meron. *Device and system for in vivo imaging*. US Patent 8,125,516. 2012. URL: <https://www.google.com/patents/US8125516>.
- [59] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain. "Wireless capsule endoscopy." In: *Nature* 405.6785 (2000), pp. 417–417. ISSN: 0028-0836. DOI: [10.1038/35013140](https://doi.org/10.1038/35013140).
- [60] *In an Iranian Image, a Missile Too Many - The New York Times*. <https://thelede.blogs.nytimes.com/2008/07/10/in-an-iranian-image-a-missile-too-many/>. (Accessed on 08/15/2017).
- [61] *Integrity of the Image report, World Press Photo*. <https://www.worldpressphoto.org/activities/research/integrity-of-the-image>. (Accessed on 08/30/2017).
- [62] *International Association of Forensic Radiographers - IAFR*. <http://www.afr.org.uk/>. (Accessed on 08/22/2017).
- [63] Hamid Jalalzadeh, Georgios F Giannakopoulos, Ferco H Berger, Judith Fronczek, Frank R W van de Goot, Udo J Reijnders, and Wietse P Zuidema. "Post-mortem imaging compared with autopsy in trauma victims—A systematic review." In: *Forensic science international* 257 (2015), pp. 29–48.
- [64] M. Jarva. *Arm structure for intra-oral x-ray device*. US Patent App. 14/006,431. 2014. URL: <http://www.google.com.pg/patents/US20140011156>.
- [65] Y. S. Jung, Y. H. Kim, D. H. Lee, and J. H. Kim. "Active Blood Detection in a High Resolution Capsule Endoscopy using Color Spectrum Transformation." In: *2008 International Conference on BioMedical Engineering and Informatics*. Vol. 1. 2008, pp. 859–862. DOI: [10.1109/BMEI.2008.216](https://doi.org/10.1109/BMEI.2008.216).

- [66] J. Kamenický et al. "PIZZARO: Forensic analysis and restoration of image and video data." In: *Forensic Science International* 264 (2016). Special Issue on the 7th European Academy of Forensic Science Conference, pp. 153–166. ISSN: 0379-0738. DOI: [10.1016/j.forsciint.2016.04.027](https://doi.org/10.1016/j.forsciint.2016.04.027).
- [67] S. Knight, S. Moschou, and M. Sorell. "Analysis of Sensor Photo Response Non-Uniformity in RAW Images." In: *e-Forensics*. Ed. by M. Sorell. Vol. 8. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer, Nov. 27, 2009, pp. 130–141.
- [68] N. Kose and J. L. Dugelay. "Classification of captured and recaptured images to detect photograph spoofing." In: *2012 International Conference on Informatics, Electronics Vision (ICIEV)*. 2012, pp. 1027–1032. DOI: [10.1109/ICIEV.2012.6317336](https://doi.org/10.1109/ICIEV.2012.6317336).
- [69] A. Kumar and Y. Zhou. "Contactless fingerprint identification using level zero features." In: *CVPR 2011 WORKSHOPS*. 2011, pp. 114–119. DOI: [10.1109/CVPRW.2011.5981823](https://doi.org/10.1109/CVPRW.2011.5981823).
- [70] Y.-G. Lee and G Yoon. "Real-Time Image Analysis of Capsule Endoscopy for Bleeding Discrimination in Embedded System Platform." In: *International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering* 5.11 (2011), pp. 583–587.
- [71] P. M. Leth. "Computerized tomography used as a routine procedure at postmortem investigations." In: *Am J Forensic Med Pathol* 30.3 (2009), pp. 219–222.
- [72] A. W. Levinovitz and N. Ringertz. *The Nobel Prize: The First 100 Years*. Imperial College Press, 2001. ISBN: 981024665X.
- [73] C.-T. Li. "Source Camera Identification Using Enhanced Sensor Pattern Noise." In: *IEEE Transactions on Information Forensics and Security* 5.2 (2010), pp. 280–287. ISSN: 1556-6013. DOI: [10.1109/TIFS.2010.2046268](https://doi.org/10.1109/TIFS.2010.2046268).
- [74] C.-T. Li, C.-Y. Chang, and Y. Li. "On the Repudiability of Device Identification and Image Integrity Verification Using Sensor Pattern Noise." In: *Information Security and Digital Forensics: First International Conference*. Ed. by D. Weerasinghe. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 19–25. ISBN: 978-3-642-11530-1. DOI: [10.1007/978-3-642-11530-1\\_3](https://doi.org/10.1007/978-3-642-11530-1_3).
- [75] C.-T. Li and Y. Li. "Color-Decoupled Photo Response Non-Uniformity for Digital Image Forensics." In: *IEEE Transactions on Circuits and Systems for Video Technology* 22.2 (2012), pp. 260–271. ISSN: 1051-8215. DOI: [10.1109/TCSVT.2011.2160750](https://doi.org/10.1109/TCSVT.2011.2160750).

- [76] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang. "Probabilistic Elastic Matching for Pose Variant Face Verification." In: *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2013, pp. 3499–3506. DOI: [10.1109/CVPR.2013.449](https://doi.org/10.1109/CVPR.2013.449).
- [77] Y. Li and C.-T. Li. "Decomposed Photo Response Non-Uniformity for Digital Forensic Analysis." In: *Forensics in Telecommunications, Information and Multimedia: Second International Conference, e-Forensics 2009, Adelaide, Australia*. Ed. by M. Sorell. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 166–172. ISBN: 978-3-642-02312-5. DOI: [10.1007/978-3-642-02312-5\\_19](https://doi.org/10.1007/978-3-642-02312-5_19).
- [78] R. B. Lindsay. *John William Strutt, 3rd Baron Rayleigh*. <https://www.britannica.com/biography/John-William-Strutt-3rd-Baron-Rayleigh>. (Accessed on 08/09/2017).
- [79] B. B. Liu, Y. Hu, and H. K. Lee. "Source camera identification from significant noise residual regions." In: *2010 IEEE International Conference on Image Processing (ICIP)*. 2010, pp. 1749–1752. DOI: [10.1109/ICIP.2010.5652426](https://doi.org/10.1109/ICIP.2010.5652426).
- [80] B. B. Liu, H. K. Lee, Y. Hu, and C. H. Choi. "On classification of source cameras: A graph based approach." In: *2010 IEEE International Workshop on Information Forensics and Security*. 2010, pp. 1–5. DOI: [10.1109/WIFS.2010.5711446](https://doi.org/10.1109/WIFS.2010.5711446).
- [81] S. Liu, B. Yang, P. C. Yuen, and G. Zhao. "A 3D Mask Face Anti-Spoofing Database with Real World Variations." In: *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2016, pp. 1551–1557. DOI: [10.1109/CVPRW.2016.193](https://doi.org/10.1109/CVPRW.2016.193).
- [82] J. Lohscheller, U. Eysholdt, H. Toy, and M. Dollinger. "Phonovibrography: Mapping High-Speed Movies of Vocal Fold Vibrations Into 2-D Diagrams for Visualizing and Analyzing the Underlying Laryngeal Dynamics." In: *IEEE Transactions on Medical Imaging* 27.3 (2008), pp. 300–309. ISSN: 0278-0062. DOI: [10.1109/TMI.2007.903690](https://doi.org/10.1109/TMI.2007.903690).
- [83] J. Lukáš, J. Fridrich, and M. Goljan. "Determining digital image origin using sensor imperfections." In: vol. 5685. 2005, pp. 249–260. DOI: [10.1117/12.587105](https://doi.org/10.1117/12.587105). URL: <http://dx.doi.org/10.1117/12.587105>.
- [84] J. Lukáš, J. Fridrich, and M. Goljan. "Digital camera identification from sensor pattern noise." In: *IEEE Transactions on Information Forensics and Security* 1.2 (2006), pp. 205–214. ISSN: 1556-6013. DOI: [10.1109/TIFS.2006.873602](https://doi.org/10.1109/TIFS.2006.873602).
- [85] J. Määttä, A. Hadid, and M. Pietikäinen. "Face spoofing detection from single images using micro-texture analysis." In: *2011 International Joint Conference on Biometrics (IJCB)*. 2011, pp. 1–7. DOI: [10.1109/IJCB.2011.6117510](https://doi.org/10.1109/IJCB.2011.6117510).

- [86] N. Mahadeo, A. Paplinski, and S. Ray. "Optimization of Iris Codes for Improved Recognition." In: *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2014, pp. 48–55. DOI: [10.1109/CVPRW.2014.13](https://doi.org/10.1109/CVPRW.2014.13).
- [87] B. Mahdian, A. Novozámský, and S. Saic. "Detecting Cyclostationarity in Re-Captured LCD Screens." In: *Journal of Forensic Research* 6.4 (2015), pp. 1–6. ISSN: 2157-7145. DOI: [10.4172/2157-7145.1000294](https://doi.org/10.4172/2157-7145.1000294).
- [88] B. Mahdian, A. Novozámský, and S. Saic. "Determination of Stop-Criterion for Incremental Methods Constructing Camera Sensor Fingerprint." In: *Digital-Forensics and Watermarking: 13th International Workshop, IWDW 2014*. Ed. by Y.-Q. Shi, H. J. Kim, F. Pérez-González, and C.-N. Yang. Taipei, Taiwan: Springer International Publishing, Oct. 2015, pp. 47–59. ISBN: 978-3-319-19321-2. DOI: [10.1007/978-3-319-19321-2\\_4](https://doi.org/10.1007/978-3-319-19321-2_4).
- [89] B. Mahdian, A. Novozámský, and S. Saic. "Identification of aliasing-based patterns in re-captured LCD screens." In: *2015 IEEE International Conference on Image Processing (ICIP)*. Québec City, Canada, Sept. 2015, pp. 616–620. DOI: [10.1109/ICIP.2015.7350872](https://doi.org/10.1109/ICIP.2015.7350872).
- [90] B. Mahdian and S. Saic. "Using noise inconsistencies for blind image forensics." In: *Image and Vision Computing* 27.10 (2009). Special Section: Computer Vision Methods for Ambient Intelligence, pp. 1497–1503. ISSN: 0262-8856. DOI: <http://dx.doi.org/10.1016/j.imavis.2009.02.001>.
- [91] B. Mahdian and S. Saic. "A bibliography on blind methods for identifying image forgery." In: *Signal Processing: Image Communication* 25.6 (2010), pp. 389–399. ISSN: 0923-5965. DOI: <http://dx.doi.org/10.1016/j.image.2010.05.003>.
- [92] C. Manfredi, L. Bocchi, S. Bianchi, N. Migali, and G. Cantarella. "Objective vocal fold vibration assessment from videokymographic images." In: *Biomedical Signal Processing and Control* 1.2 (2006). Voice Models and Analysis for Biomedical Applications, pp. 129–136. ISSN: 1746-8094. DOI: <https://doi.org/10.1016/j.bspc.2006.06.001>. URL: <http://www.sciencedirect.com/science/article/pii/S1746809406000243>.
- [93] K. Matsushita and H. Kitazawa. "An improved camera identification method based on the texture complexity and the image restoration." In: *ICHIT*. Ed. by G. Lee, D. Howard, J. J. Kang, D. Slezak, T. N. Ahn, and C.-H. Yang. Vol. 321. ACM International Conference Proceeding Series. ACM, 2009, pp. 171–175. ISBN: 978-1-60558-662-5.

- [94] S. McCloskey. "Confidence weighting for sensor fingerprinting." In: *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2008, pp. 1–6. DOI: [10.1109/CVPRW.2008.4562986](https://doi.org/10.1109/CVPRW.2008.4562986).
- [95] A. Moglia, A. Menciassi, and P. Dario. "Recent Patents on Wireless Capsule Endoscopy." In: *Recent Patents on Biomedical Engineering* 1.1 (2008), pp. 24–33. ISSN: 2211-3320/1874-7647. DOI: [10.2174/1874764710801010024](https://doi.org/10.2174/1874764710801010024).
- [96] *Multimedia Forensic Investigations*. <https://revealproject.eu/multimedia-forensic-investigations/>. (Accessed on 08/15/2017).
- [97] A. Novozámský. "Source Camera Identification Based on PRNU Invariant to Zoom." In: *Doktorandské dny 2011 sborník workshopu doktorandů FJFI oboru Matematické inženýrství*. Prague, Czechia, Nov. 2011, pp. 163–173. URL: <http://kmwww.fjfi.cvut.cz/ddny/historie/11-sbornik.pdf>.
- [98] A. Novozámský. "Rima Glottidis Segmentation by Thresholding Using Graph Cuts." In: *Doktorandské dny 2012 sborník workshopu doktorandů FJFI oboru Matematické inženýrství*. Prague, Czechia, Nov. 2012, pp. 177–185. URL: <http://kmwww.fjfi.cvut.cz/ddny/historie/12-sbornik.pdf>.
- [99] A. Novozámský, B. Mahdian, and S. Saic. "Source camera identification invariant to zoom." In: *International Computer Vision Summer School 2012*. Sicily, Italy, July 2012. URL: <http://iplab.dmi.unict.it/icvss2012/posters.html>.
- [100] A. Novozámský and M. Šorel. "Copy-move forgery detection using JPEG compression model." In: *Submitted to Forensic Science International* (2017).
- [101] A. Novozámský, J. Sedlář, A. Zita, J. G. Švec, B. Zitová, J. Flusser, and D. Hauzar. "Computer-Assisted Evaluation of Videokymographic Data." In: *EFMI STC Prague Data and Knowledge for Medical Decision Support*. Prague, Czechia, Apr. 2013.
- [102] A. Novozámský, J. Sedlář, A. Zita, Ch. Herbst, J. G. Švec, B. Zitová, and J. Flusser. "VKFD: Computerized Analysis of Videokymographic Data." In: *10. PEVOC - Pan European Voice Conference*. Prague, Czechia, Aug. 2013.
- [103] A. Novozámský, J. Sedlář, A. Zita, F. Šroubek, J. Flusser, J. G. Švec, J. Vydrová, and B. Zitová. "Image analysis of videokymographic data." In: *2015 IEEE International Conference on Image Processing (ICIP)*. Québec City, Canada, Sept. 2015, pp. 78–82. DOI: [10.1109/ICIP.2015.7350763](https://doi.org/10.1109/ICIP.2015.7350763).

- [104] A. Novozámský, J. Flusser, I. Tachecí, L. Sulík, J. Bureš, and O. Krejcar. "Automatic blood detection in capsule endoscopy video." In: *Journal of Biomedical Optics* 21.12 (2016), p. 126007. DOI: [10.1117/1.JBO.21.12.126007](https://doi.org/10.1117/1.JBO.21.12.126007).
- [105] A. L. S. Orozco, D. M. A. González, J. R. Corripio, L. J. G. Villalba, and J. C. Hernandez-Castro. "Source identification for mobile devices, based on wavelet transforms combined with sensor imperfections." In: *Computing* 96.9 (2014), pp. 829–841. ISSN: 1436-5057. DOI: [10.1007/s00607-013-0313-5](https://doi.org/10.1007/s00607-013-0313-5).
- [106] G. Pan, L. Sun, Z. Wu, and S. Lao. "Eyeblink-based Anti-Spoofing in Face Recognition from a Generic Webcam." In: *2007 IEEE 11th International Conference on Computer Vision*. 2007, pp. 1–8. DOI: [10.1109/ICCV.2007.4409068](https://doi.org/10.1109/ICCV.2007.4409068).
- [107] S. C. Park et al. "Sensitivity of the suspected blood indicator: An experimental study." In: *World J Gastroenterol* 18.31 (2012). 22919250[pmid], pp. 4169–4174. ISSN: 1007-9327. DOI: [10.3748/wjg.v18.i31.4169](https://doi.org/10.3748/wjg.v18.i31.4169).
- [108] B. Penna, T. Tillo, M. Grangetto, E. Magli, and G. Olmo. "A technique for blood detection in wireless capsule endoscopy images." In: *Signal Processing Conference, 2009 17th European*. 2009, pp. 1864–1868.
- [109] M. Pennazio. "Small-Bowel Endoscopy." In: *Endoscopy* 36.01 (2004). 32, pp. 32–41. ISSN: 0013-726X. DOI: [10.1055/s-2004-814116](https://doi.org/10.1055/s-2004-814116).
- [110] M. Pennazio. "Capsule endoscopy: Where are we after 6 years of clinical use?" In: *Digestive and Liver Disease* 38.12 (2006), pp. 867–878. ISSN: 1590-8658. DOI: <http://dx.doi.org/10.1016/j.dld.2006.09.007>.
- [111] T. de F. Pereira, J. Komulainen, A. Anjos, J. M. de Martino, A. Hadid, M. Pietikäinen, and S. Marcel. "Face liveness detection using dynamic texture." In: *EURASIP Journal on Image and Video Processing* 2014.1 (2014), p. 2. ISSN: 1687-5281. DOI: [10.1186/1687-5281-2014-2](https://doi.org/10.1186/1687-5281-2014-2).
- [112] *Photo Tampering throughout History*. [http://pth.izitru.com/1864\\_13\\_00.html](http://pth.izitru.com/1864_13_00.html). (Accessed on 08/28/2017).
- [113] J. L. Prince and J. Links. *Medical Imaging Signals and Systems (2nd Edition)*. Pearson, 2014. ISBN: 0132145189.
- [114] Q. Qiu, H. K. Schutte, L. Gu, and Q. Yu. "An automatic method to quantify the vibration properties of human vocal folds via videokymography." In: *Folia Phoniatr Logop* 55.3 (2003), pp. 128–136.
- [115] J. W. S. B. Rayleigh. *The Theory of Sound*. The Theory of Sound sv. 1. Macmillan, 1877. URL: <https://books.google.cz/books?id=hd8EAAAAYAAJ>.

- [116] W. C. Röntgen. "On a new kind of rays." In: *Science* 3.59 (1896), pp. 227–231. ISSN: 0036-8075. DOI: [10.1126/science.3.59.227](https://doi.org/10.1126/science.3.59.227).
- [117] C. A. Rosen. "Stroboscopy as a research instrument: development of a perceptual evaluation tool." In: *Laryngoscope* 115.3 (2005), pp. 423–428.
- [118] K. Rosenfeld and H. T. Sencar. "A study of the robustness of PRNU-based camera identification." In: *Media Forensics and Security*. Ed. by E. J. Delp, J. Dittmann, N. D. Memon, and P. W. Wong. Vol. 7254. SPIE Proceedings. SPIE, 2009, p. 72540.
- [119] *Safe Root Canal Treatment*. <http://www.rootcanaltruth.com/it-works.html>. (Accessed on 08/11/2017).
- [120] J. Sedlář, A. Novozámský, J. G. Švec, B. Zitová, and J. Flusser. "Measurement of Vocal Fold Features in Videokymography Images." In: *BioImage Informatics*. Dresden, Germany, Sept. 2012.
- [121] S. K. Shah, J. K. Lee, and M. E. Celebi. "Classification of Bleeding Images in Wireless Capsule Endoscopy using HSI Color Domain and Region Segmentation." In: *proceeding of 2007 New England American Society for Engineering Education Conference*. 2007.
- [122] X. Shen, Z. Lin, J. Brandt, and Y. Wu. "Detecting and Aligning Faces by Image Retrieval." In: *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2013, pp. 3460–3467. DOI: [10.1109/CVPR.2013.444](https://doi.org/10.1109/CVPR.2013.444).
- [123] E. Silva, T. Carvalho, A. Ferreira, and A. Rocha. "Going deeper into copy-move forgery detection: Exploring image telltales via multi-scale analysis and voting processes." In: *Journal of Visual Communication and Image Representation* 29 (2015), pp. 16–32. ISSN: 1047-3203. DOI: <http://dx.doi.org/10.1016/j.jvcir.2015.01.016>.
- [124] M. Steinebach, H. Liu, P. Fan, and S. Katzenbeisser. "Cell phone camera ballistics: attacks and countermeasures." In: *Proc. SPIE*. Vol. 7542. 2010, 75420B–75420B–9. DOI: [10.1117/12.838870](https://doi.org/10.1117/12.838870).
- [125] J. G. Svec and H. K. Schutte. "Videokymography: high-speed line scanning of vocal fold vibration." In: *J Voice* 10.2 (1996), pp. 201–205.
- [126] M. M. Ter-Pogossian, M. E. Phelps, E. J. Hoffman, and N. A. Mullani. "A Positron-Emission Transaxial Tomograph for Nuclear Imaging (PETT)." In: *Radiology* 114.1 (1975), pp. 89–98. DOI: [10.1148/114.1.89](https://doi.org/10.1148/114.1.89).

- [127] S. Thavalengal, P. Bigioi, and P. Corcoran. "Evaluation of combined visible/NIR camera for iris authentication on smartphones." In: *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2015, pp. 42–49. DOI: [10.1109/CVPRW.2015.7301318](https://doi.org/10.1109/CVPRW.2015.7301318).
- [128] *The International Society for Forensic Radiology and Imaging (IS-FRI)*. <http://www.isfri.org/>. (Accessed on 08/22/2017).
- [129] *The Nobel Prize in Physiology or Medicine for 2003*. [https://www.nobelprize.org/nobel\\_prizes/medicine/laureates/2003/press.html](https://www.nobelprize.org/nobel_prizes/medicine/laureates/2003/press.html). (Accessed on 08/09/2017).
- [130] *The Offside Blog*. <https://offsidesportsphotography.wordpress.com/>. (Accessed on 08/24/2017).
- [131] T. Thongkamwitoon, H. Muammar, and P. L. Dragotti. "Robust image recapture detection using a K-SVD learning approach to train dictionaries of edge profiles." In: *2014 IEEE International Conference on Image Processing (ICIP)*. 2014, pp. 5317–5321. DOI: [10.1109/ICIP.2014.7026076](https://doi.org/10.1109/ICIP.2014.7026076).
- [132] T. Thongkamwitoon, H. Muammar, and P. L. Dragotti. "An Image Recapture Detection Algorithm Based on Learning Dictionaries of Edge Profiles." In: *IEEE Transactions on Information Forensics and Security* 10.5 (2015), pp. 953–968. ISSN: 1556-6013. DOI: [10.1109/TIFS.2015.2392566](https://doi.org/10.1109/TIFS.2015.2392566).
- [133] D. Tralic, I. Zupancic, S. Grgic, and M. Grgic. "CoMoFoD-New database for copy-move forgery detection." In: *Proceedings ELMAR-2013*. 2013, pp. 49–54.
- [134] M. Visentini-Scarzanella and P. L. Dragotti. "Video jitter analysis for automatic bootleg detection." In: *2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*. 2012, pp. 101–106. DOI: [10.1109/MMSP.2012.6343423](https://doi.org/10.1109/MMSP.2012.6343423).
- [135] N. B. A. Warif, M. Y. I. Idris, A. W. A. Wahab, and R. Salleh. "An evaluation of Error Level Analysis in image forensics." In: *2015 5th IEEE International Conference on System Engineering and Technology (ICSET)*. 2015, pp. 23–28. DOI: [10.1109/ICSEngT.2015.7412439](https://doi.org/10.1109/ICSEngT.2015.7412439).
- [136] W. E. Whitehead. *Gastrointestinal Motility Disorders of the Small Intestine, Large Intestine, Rectum, and Pelvic Floor*. IFFGD Fact Sheet No. 162. 2001.
- [137] G. Wu, X. Kang, and K. J. R. Liu. *A Context Adaptive Predictor of Sensor Pattern Noise for Camera Source Identification*. Tech. rep. 2012.
- [138] D. Yi, Z. Lei, and S. Z. Li. "Towards Pose Robust Face Recognition." In: *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2013, pp. 3539–3545. DOI: [10.1109/CVPR.2013.454](https://doi.org/10.1109/CVPR.2013.454).

- [139] Y. Yuan, B. Li, and M. Q. H. Meng. "Bleeding Frame and Region Detection in the Wireless Capsule Endoscopy Video." In: *IEEE Journal of Biomedical and Health Informatics* 20.2 (2016), pp. 624–630. ISSN: 2168-2194. DOI: [10.1109/JBHI.2015.2399502](https://doi.org/10.1109/JBHI.2015.2399502).
- [140] C. Zhang and H. Zhang. "Identifying color image origin using curvelet transform." In: *Proc. 17th IEEE Int Image Processing (ICIP) Conf.* 2010, pp. 2125–2128.
- [141] G. Zhao and M. Pietikäinen. "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions." In: *IEEE Trans. Pattern Anal. Mach. Intell.* 29.6 (June 2007), pp. 915–928. ISSN: 0162-8828. DOI: [10.1109/TPAMI.2007.1110](https://doi.org/10.1109/TPAMI.2007.1110).
- [142] B. Zitová, F. Šroubek, J. Kamenický, B. Mahdian, A. Novozámský, S. Saic, A. Zita, Z. Šíma, P. Švarc, and J. Hořínek. "Imaging device identification and detection of image tampering." In: *Abstract book of 7th European Academy of Forensic Science Conference*. Prague, Czechia, Sept. 2015.

Part III

ATTACHED ARTICLES





## PIZZARO: Forensic analysis and restoration of image and video data



Jan Kamenicky<sup>a</sup>, Michal Bartos<sup>a</sup>, Jan Flusser<sup>a</sup>, Babak Mahdian<sup>a</sup>, Jan Kotera<sup>a</sup>,  
Adam Novozamsky<sup>a</sup>, Stanislav Saic<sup>a</sup>, Filip Sroubek<sup>a</sup>, Michal Sorel<sup>a</sup>, Ales Zita<sup>a</sup>,  
Barbara Zitova<sup>a,\*</sup>, Zdenek Sima<sup>b</sup>, Petr Svarc<sup>b</sup>, Jan Horinek<sup>c</sup>

<sup>a</sup> Institute of Information Theory and Automation, The Czech Academy of Sciences, Pod vodárenskou veží 4, Prague, Czech Republic

<sup>b</sup> Institute of Criminalistics, Audio-Video Department, Prague, Czech Republic

<sup>c</sup> National Drug Headquarters of the Criminal Police and Investigation Service of the Police of the Czech Republic, Informatics Department, Prague, Czech Republic

## ARTICLE INFO

## Article history:

Available online 28 April 2016

## Keywords:

Image forensic analysis  
Image restoration  
Image tampering detection  
Image source identification

## ABSTRACT

This paper introduces a set of methods for image and video forensic analysis. They were designed to help to assess image and video credibility and origin and to restore and increase image quality by diminishing unwanted blur, noise, and other possible artifacts. The motivation came from the best practices used in the criminal investigation utilizing images and/or videos. The determination of the image source, the verification of the image content, and image restoration were identified as the most important issues of which automation can facilitate criminalists work. Novel theoretical results complemented with existing approaches (LCD re-capture detection and denoising) were implemented in the PIZZARO software tool, which consists of the image processing functionality as well as of reporting and archiving functions to ensure the repeatability of image analysis procedures and thus fulfills formal aspects of the image/video analysis work. Comparison of new proposed methods with the state of the art approaches is shown. Real use cases are presented, which illustrate the functionality of the developed methods and demonstrate their applicability in different situations. The use cases as well as the method design were solved in tight cooperation of scientists from the Institute of Criminalistics, National Drug Headquarters of the Criminal Police and Investigation Service of the Police of the Czech Republic, and image processing experts from the Czech Academy of Sciences.

© 2016 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

In our world, digital images are an extensively used medium of communication that compactly and efficiently conveys a huge amount of information about our surrounding. However two important questions should be considered – how much we can trust all these photographs which are not necessarily obtained from a trustworthy source and do we notice everything contained in images, which are often acquired under non-ideal, inadequate conditions and thus possibly blurred, noisy, or with other types of unwanted artifacts.

Our research addressed two related topics – the quality of image data with respect to their credibility and origin, and image restoration aimed at diminishing unwanted blur, noise, and other possible artifacts. The motivation came from the best practices used in the criminal investigation utilizing images and/or videos. The *determination of the image source*, the *verification of the image content*, and possible *image restoration* were identified as important issues, of which automation can notably facilitate criminalist's work (see Fig. 1). Situations, when the origin of images (child pornography) or their authenticity (insurance frauds) is questioned, occur with increasing frequency nowadays.

Regarding the image source determination, the ability to link individual pictures to appropriate acquisition devices (camera, scanner, cell phone, etc.) can help investigators for example to indicate a particular camera type, or to detect cases, when a digital photo has been re-captured from an LCD screen. For the particular camera unit identification and model specification we have applied approaches based on photo response non uniformity (PRNU) and JPEG quantization tables, respectively. PRNU is known to be able to describe pixel sensitivity to light and is present in the sensor

\* Corresponding author.

E-mail addresses: [kamenik@utia.cas.cz](mailto:kamenik@utia.cas.cz) (J. Kamenicky), [bartos@utia.cas.cz](mailto:bartos@utia.cas.cz) (M. Bartos), [flusser@utia.cas.cz](mailto:flusser@utia.cas.cz) (J. Flusser), [mahdian@utia.cas.cz](mailto:mahdian@utia.cas.cz) (B. Mahdian), [kotera@utia.cas.cz](mailto:kotera@utia.cas.cz) (J. Kotera), [novozamsky@utia.cas.cz](mailto:novozamsky@utia.cas.cz) (A. Novozamsky), [ssaic@utia.cas.cz](mailto:ssaic@utia.cas.cz) (S. Saic), [sroubekf@utia.cas.cz](mailto:sroubekf@utia.cas.cz) (F. Sroubek), [sorel@utia.cas.cz](mailto:sorel@utia.cas.cz) (M. Sorel), [zita@utia.cas.cz](mailto:zita@utia.cas.cz) (A. Zita), [zitova@utia.cas.cz](mailto:zitova@utia.cas.cz) (B. Zitova), [zdenek.sima1@pcr.cz](mailto:zdenek.sima1@pcr.cz) (Z. Sima), [pizzaro@utia.cas.cz](mailto:pizzaro@utia.cas.cz) (P. Svarc), [jan.horinek@seznam.cz](mailto:jan.horinek@seznam.cz) (J. Horinek).

<http://dx.doi.org/10.1016/j.forensiint.2016.04.027>

0379-0738/© 2016 Elsevier Ireland Ltd. All rights reserved.

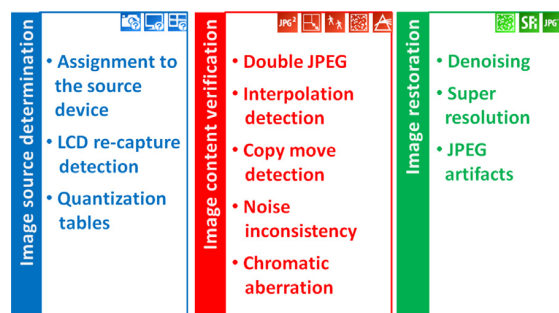


Fig. 1. An overview of the developed and implemented tools for image forensic analysis in PIZZARO.

pattern noise. The detection of LCD screen acquisition is done using structural and color abnormalities in recaptured images. This task is challenging due to advances in image capture technologies. It is possible to re-capture high quality images in such a way they are hard to be distinguished from the originals by an untrained eye.

Various approaches developed for detection of tampering and forgery traces in image content range from a simple metadata analysis to a complex shadow and scene lighting analysis using mathematical models. The simple metadata analysis often provides only a weak indicator of forgery and usually is not applicable in real cases. On the other hand, the manual scene analysis exploiting knowledge of scene lighting and/or geometry parameters creates strong forgery indicators but usually requires complex tuning for every case by professionals with deep expertise in the field. We have provided a set of methods that can help investigators in their decision making. These methods belong to the family of passive (blind) image forensic methods, which does not require any prior information about images under verification.

Specifically, we focused on methods for detecting double compression, interpolation, copy–paste action, and noise and chromatic aberration inconsistencies. Double compression is introduced to an image while altering the JPEG image in a photo-editing software, when the image is decompressed and after the manipulations are carried out the image is re-saved and thus compressed again. Interpolation, required when the image is geometrically transformed (i.e., rotation, scaling, skewing), is often applied on the image to be able to merge two or more images together to create high quality and consistent image forgeries. In copy–move forgeries, a part of the image is copied and pasted into another part of the same image, with the intention to hide an object or a region of the image. Methods for copy–move forgery detection search for similar overlapping image blocks, with possibly blurred borders (to achieve smooth transition between the original image and pasted blocks). Noise inconsistencies are often introduced while concealing the tampering traces, when locally random noise is added to the forged image regions. Finally, inconsistencies in chromatic aberrations across the image can indicate an image forgery, too. The chromatic aberrations are introduced to the image due to inability of a camera lens to focus all colors to the same convergence point. It is manifested as a regular pattern of slight blur and/or color shifts. Irregularities in this pattern could imply image tampering.

Finally, the last part of our work is related to improving image quality by reducing noise and blur. A paradigm for forensic specialists examining images and videos from surveillance cameras is to deal with cases that often exhibit insufficient resolution and compression artifacts, which prevents correct identification of subjects, such as human faces or car license plates. Moreover, omnipresent image noise is a frequent image degradation, typically visible in images taken in low-light conditions. We design a methodology for removing noise and compression

artifacts and even increasing data resolution, if applicable (e.g., several frames from a video sequence are available).

Proposed work is based on tight cooperation of scientists from the Institute of Criminalistics, National Drug Headquarters of the Criminal Police and Investigation Service of the Police of the Czech Republic, and image processing experts from the Czech Academy of Sciences. All novel theoretical results together with existing methods for LCD image re-capture and denoising, were implemented in the PIZZARO software [1] to provide required tools. In addition to the image processing functionality, we also paid attention to formal aspects of the image/video analysis, which play important role in the forensic practice. The PIZZARO software enables to create log files and reports to ensure the repeatability of image analysis procedures, when all taken steps with parameter settings and achieved outputs are recorded for future use/verification together with the operator identity and archiving information such as the file paths and dates.

After the review of other existing software solutions in Section 2 we will introduce methods proposed for solving the above mentioned three issues – source device analysis in Section 3, image content verification in Section 4, and image restoration in Section 5. Section 6 introduces main features of the resulting software package PIZZARO (<http://pizzaro.utia.cas.cz>), implementing proposed algorithms. Illustrative use cases with examples of the PIZZARO application are described in Section 7 and Section 8 concludes the paper.

## 2. Existing software solutions

Numerous software packages have been introduced in recent years for verifying the integrity of digital images. Some of them utilized sophisticated theoretical and scientific approaches and some others rather focused on practical and ad-hoc methods. In general, Forensic Toolkit (FTK) [2] has become the most popular solution in the digital forensic field and it has been widely used by digital forensic investigators. Fourandsix [3] has introduced an image forensic tool providing image authentication functionalities by using JPEG signatures. This software is mainly oriented on media publishers and forensic investigators. Verifeyed [4] has developed software focused on detecting manipulated photos and PDF files for insurance and corporate security markets. Belkasoft [5] introduced forensic software that enables to analyze the validity of digital images by using a set of separate basic image forensic methods. They also focused on insurance market. Forensic Pathways [6] has developed and introduced to the market an image ballistics software that aims to meet main needs of forensic investigators in the field. Last, but not least, we have to mention the Image Error Level Analyser [7] which is a free and easily accessible image forensic tool that has become very popular among the community despite its limitations and high false positive rates.

Also the area of image and video restoration and enhancement is experiencing a lot of new software packages. For example, we can mention Amped Five [8] (started as MIPE [9]), that developed an image and video enhancement (super-resolution) software. This software also provides image forensic and ballistics tools targeting the digital forensic experts. Another tool that can be mentioned is MotionDSP [10] that provides real-time super-resolution and video enhancement techniques. They also deliver GPU accelerated versions of their software. Cognitech [11] provides methods enhancing quality and resolution of (mainly facial) videos for the forensic market. Another software package providing super resolution, denoising, and deblurring has been developed by RTCVision [12]. Impress [13] is another company providing video enhancing tools accelerated by utilizing GPUs. The authors of the webpage [14] provide several alternative SR toolboxes.

Despite the fact that there are numerous image and video forensic and enhancement software packages available on market, we can note that none of them provides complete solution.

Inspecting and enhancing digital images and videos is complex process and often a combination of different methods and solutions is needed to derive meaningful and desired output.

### 3. Image source determination

Methods assigning a digital image to the source camera are typically based on the fact that image sensors suffer from several fundamental and technology related imperfections resulting in their performance limitations and noise. For example, if we take a picture of an absolutely evenly lit scene, the resulting digital image usually still exhibits small changes in intensity among individual pixels which is partly because of pattern noise, readout noise or shot noise. While readout noise or shot noise are random components, the pattern noise is deterministic and remain approximately the same if multiple pictures of the same scene are taken. As a result, pattern noise can be used as the sensor fingerprint of cameras.

Alternatively, if the digital image is the result of LCD recapturing (i.e., a photo displayed on an LCD is recaptured by camera), often we can observe periodic detectable artifacts in the recaptured photo. The last method for analysis of the image acquisition device is based on the evaluation of the camera and image attributes related to their resolution and the way how they were preprocessed.

#### 3.1. Assignment to the source device

It has been shown that photo response non uniformity (PRNU, describing the pixel sensitivity to light) has a dominant presence in the sensor pattern noise [15,16] and can be utilized as the sensor fingerprint because of its stability and discrimination power. Most of the state-of-the-art source identification methods are based on the method proposed by Jessica Fridrich et al. (e.g., [15,17]). Here, the image acquisition process is modeled in the following way:

$$g = u + u \cdot \Gamma + Y, \quad (1)$$

where  $g$  denotes the image produced by the camera,  $u$  denotes the noise-free image (the perfect image of the scene),  $\Gamma$  denotes the PRNU noise and  $Y$  stands for all other additive or negligible noise components. We have proposed a modification of the standard approach, addressing the problems with the time demanding PRNU estimation and with the situation when the images are taken with various zoom.

For a given camera, PRNU noise can be estimated by averaging multiple images  $g_i$ ,  $i = 1, \dots, N$  captured by this camera. Prior to averaging, the scene content is suppressed from the image. This preprocessing reduces the required number of images that is needed to achieve a robust PRNU estimation. The suppression is realized by application of a denoising filter  $\mathcal{F}$  and averaging the noise residuals instead. We will denote these residuals by  $\hat{g}_i$  (i.e.,  $\hat{g}_i = g_i - \mathcal{F}(g_i)$ ). Hence, the deterministic components of the camera sensor noise are computed in the following way:

$$\Gamma_{\text{sensor}} = \frac{1}{N} \sum_{i=1}^N \hat{g}_i = \frac{1}{N} \sum_{i=1}^N (g_i - \mathcal{F}(g_i)). \quad (2)$$

Alternatively, maximum likelihood estimation (MLE) instead of simple averaging is often employed.

To reduce the false positive rate, sensor fingerprints are enhanced by Wiener filtering in the frequency domain (e.g., to reduce JPEG compression artifacts) as well as by using a linear pattern removal process through zero-mean operation (e.g., to remove traces of color filter array – CFA – interpolation) [16]. To summarize,  $\Gamma_{\text{sensor}}$  is denoted as the basic version of the sensor fingerprint of the camera.

The assignment test of a digital image to the possible source camera has been carried out by evaluating the similarity measure of two sensor fingerprints. One is obtained from the image under investigation and second from the set of camera reference images. Having available two different sensor fingerprints  $\Gamma_{s_1}$  and  $\Gamma_{s_2}$ , we can measure their similarity by employing a normalized correlation:

$$\text{corr}(\Gamma_{s_1}, \Gamma_{s_2}) = \frac{\langle (\Gamma_{s_1} - \bar{\Gamma}_{s_1}), (\Gamma_{s_2} - \bar{\Gamma}_{s_2}) \rangle}{\|\Gamma_{s_1} - \bar{\Gamma}_{s_1}\| \|\Gamma_{s_2} - \bar{\Gamma}_{s_2}\|}, \quad (3)$$

where  $\bar{X}$  denotes mean of the vector  $X$ ,  $\langle \rangle$  denotes inner product and  $\|\cdot\|$  denotes  $L_2$  norm. If the calculated correlation is above a certain threshold, we accept the hypothesis that the tested image has been acquired by the camera.

The implemented method differs from the one described in [15,18]. It only uses the central portions of images which leads to more robust results when images are strongly impacted by vignetting. The vignetting is typically more profound at non-central image parts thus filtering out outer parts of images increases the probability of achieving higher true positive (TP) rate.

In Fig. 2 we demonstrate on 43 images taken with different optical zoom how the correlation values increase if the central parts (bottom graph) are used instead of the full-size images (upper graph). The two histograms in Fig. 3 show that the correlations of the reference fingerprint with non-corresponding cameras are below the values with the corresponding camera even when the central part approach was applied, so the cropping does not cause increase in the number of false positives. The minimum correlation of the

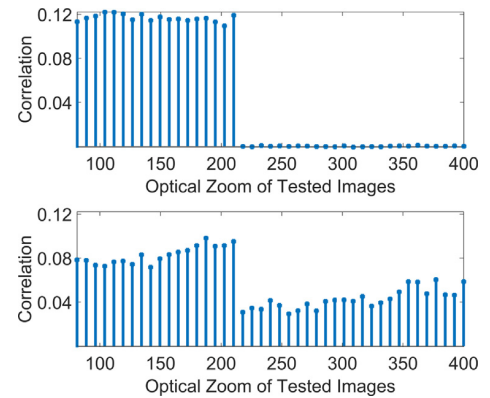


Fig. 2. Correlation values for images taken by the reference camera but with an increasing zoom; the  $\Gamma_{\text{sensor}}$  was created from images taken with the zoom 50. Top: using the full-size image approach; bottom: using the central part approach.

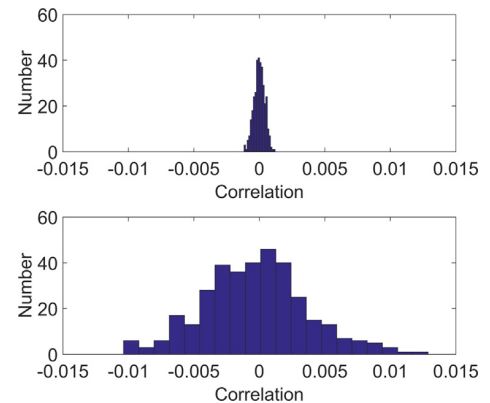


Fig. 3. Histogram of the correlation values from the false positive rate testing (10 different cameras). Top: using the full-size image approach; bottom: using the central part approach.

true test (Fig. 2 bottom) is more then 2 times higher then the maximum of false test in Fig. 3 bottom. Moreover, the central portion approach significantly reduce the computational time.

To increase the robustness of the correct identification we recommend to create several  $\Gamma_{\text{sensor}}$  with different zoom. It eliminates the need for the central part. Our software includes this feature that multiple  $\Gamma_{\text{sensor}}$  can be associated with a single camera.

### 3.2. LCD image re-capture

In order to be able to identify images which were created by re-capturing a photograph from a LCD screen, we have implemented three types of features which are able to reflect the abnormalities of recaptured image [19]. They are used together with a pre-trained support vector machine (SVM) classifier. Design of the feature set is focused on detection of certain image aberrations occurring during the re-capture process.

On LCD-originating images periodic texture pattern can be often observed. This is due to the aggregation of regular structures of LCD. Even though this pattern can be partially avoided by marginal image blurring or by higher quality data acquisition, related loss of fine details is inevitable and we exploit this fact in the first two types of features. The third type of features is based on color characteristic. The color space of the resulting recaptured image is usually less saturated comparing to the original due to the generally smaller color gamut of LCD screens than a typical camera.

The set of implemented features consists of:

- **LBP features** – 80 features are evaluated using the local binary patterns (LBP) operator  $\text{LBP}_{P,R}^{\text{riu2}}$  with  $P = \{8, 16, 24, 24\}$  and  $R = \{1, 2, 3, 4\}$ , respectively [20]. It is defined by means of a set of  $P$  neighbor pixels for the given pixel, equally distributed on a circle of radius  $R$  (see Fig. 4). To achieve gray scale and contrast invariance the operator is defined as

$$\text{LBP}_{P,R} = \sum_{p=0}^{P-1} H(g(p) - g(c)) 2^p, \quad (4)$$

where  $g(c)$  is the intensity of the central pixel,  $g(p)$  is the intensity of the  $p$ th neighbor and  $H(\cdot)$  is Heaviside step function. Note, that by using  $2^P$  binomial factor, each spatial structure is assigned unique  $\text{LBP}_{P,R}$  number. To achieve rotation invariance:

$$\text{LBP}_{P,R}^{\text{riu2}} = \begin{cases} \sum_{p=0}^{P-1} H(g(p) - g(c)) & U_{P,R} \leq 2 \\ P + 1 & \text{otherwise,} \end{cases} \quad (5)$$

where  $U_{P,R}$  is the uniformity operator, which corresponds to number of  $\text{LBP}_{P,R}$  spatial transitions  $U_{P,R} = |H(g(P-1) - g(c)) - H(g(0) - g(c))| + \sum_{p=1}^{P-1} |H(g(p) - g(c)) - H(g(p-1) - g(c))|$ .

In this way, only the most uniform neighbor sets are taken into consideration.

- **Multi-scale wavelet statistics** – the first two geometric moments (the mean and the variance) of all the high-pass bands from the Haar wavelet decomposition up to the level 3, computed separately for R, G, and B channels – 54 features in total.
- **Color features** – Average pixel value of each color channel; RG, RB, and BG correlations between color channels; center of mass of histogram  $h(i)$  of each color channel, where  $h(i)$  is the number of pixels with intensities  $i - 1$  or  $i + 1$ ; RGB pairs energy ratios  $E_1 = |G|^2/|B|^2$ ,  $E_2 = |G|^2/|R|^2$ ,  $E_3 = |B|^2/|R|^2$  [21]; mean, standard deviation and skewness, computed for the H, S, and V color channels, respectively [22].

The implemented SVM classifier is trained using the described features on a set of original images with their recaptured counterparts and then used for verification. We have used a database of 12,000 images, half of which were recaptured from numerous LCD/camera combinations. 2000 images were manually recaptured from our original photographs, whereas the rest were general images taken from Internet sources.

The recommended usage of the proposed approach starts with the classifier training using the samples from camera/LCD set in question. As is apparent from practice the information about the camera and LCD devices (e.g., devices owned by forgery suspect) is often available during the investigation.

### 3.3. Quantization tables

The last method aims to identify the set of camera models possibly used for the acquisition of the analyzed image. The proposed method is based on extraction of a certain set of features from the digital image file and matching them to a set of camera models. For instance, having a digital image of resolution of  $1000 \times 800$  pixels and a claim that the digital image has been captured by a particular camera model (camera model name is stored in digital image metadata) and not been modified since, we can check if that particular camera device can produce digital images with such resolution. If we know that the camera produces only digital images with resolutions  $2592 \times 1944$ ,  $1600 \times 1200$ , and  $640 \times 480$ , we obviously can draw a conclusion that the above claim is false and the digital image was not produced by this camera. Otherwise, the tested model can be a potential source of the image.

The above example illustrates how the image resolution can be employed as a feature. Numerous similar features can be used for the image-camera matching process [23]. More formally, we say that a digital image has attributes, and the image metadata are their respective values, which characterize the digital image. Essentially, some attribute values are dependent on the camera with which the digital image has been taken. We refer to such camera associated features stored in the digital image as camera fingerprints. The properties that can explicitly characterize an acquisition device (camera) include, e.g., its producer and model,

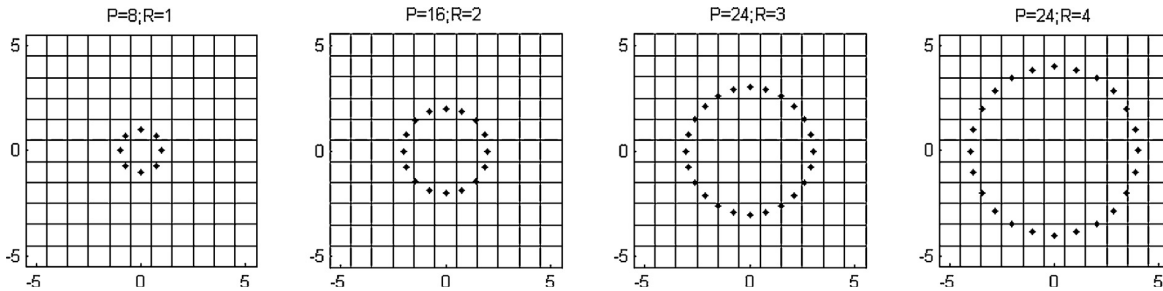


Fig. 4. Circularly symmetric neighbor sets for  $\text{LBP}_{8,1}^{\text{riu2}}$ ,  $\text{LBP}_{16,2}^{\text{riu2}}$ ,  $\text{LBP}_{24,3}^{\text{riu2}}$  and  $\text{LBP}_{24,4}^{\text{riu2}}$ .

the output file format, imaging sensor properties, the digital zoom interpolation method, the color filter array interpolation method used to encode an image, etc. Some of these properties identify a camera uniquely, and some of them can be considered as camera model fingerprints. In the provided method, quantization tables (QTs), which encode digital images to JPEG format are used to match photos to their potential source camera models.

We assume a fixed tuple of properties sufficient for unique identification of any camera. We denote such a tuple by  $c\vec{m}$ , a camera ID vector. Next, we assume a tuple of camera attributes, whose values pose suitable fingerprints of available cameras in our database. We denote such a tuple of fingerprints by  $\vec{\theta}$ , a fingerprint vector.

Specifically, given a “testing” tuple  $(c\vec{m}_0, \vec{\theta})$ , our default position is that  $\vec{\theta}$  cannot be a real fingerprint vector of  $c\vec{m}_0$ . Accordingly, we set out the following null hypothesis:

$$H_0 : \text{“}\vec{\theta} \text{ cannot be a fingerprint of } c\vec{m}_0\text{”}. \quad (6)$$

To accept or reject this hypothesis, we downloaded over one million digital images from a typical photo sharing site and extracted a reference fingerprint data set. To discard non-original (i.e., manipulated) images and create a reliable reference data set, photos containing obvious traces of modifications were eliminated. To further eliminate non-original images, only those that form sufficiently big clusters of images with the same paired producer model and their QTs are retained and employed to accept or reject the null hypothesis – to identify a set of potential source camera models for given tested image. The classification is based on the luminance QTs only and a threshold-based test replaced originally used statistical testing.

#### 4. Image content verification

Digital forensic methods for detection of forgery traces can be classified into two main categories – data hiding approaches and methods working with digital signatures. By data hiding we refer to methods embedding secondary data into the image such as digital watermarking approaches do [24]. Digital watermarking assumes an inserting of a digital watermark at the source side (e.g., camera) and verifying the mark integrity at the detection side. Watermarks are mostly inseparable from the digital image they are embedded in and they undergo the same transformations as the image itself. The latter is their main advantage – it is difficult to overcome them – but at the same time their disadvantage – they have to be designed robustly not to be degraded by expected user operations.

Methods using the digital signatures are based on extracting unique features from the image at the source side and encoding these features into the so-called digital signature. These signatures are then used to verify the image integrity. This research direction is popular thus numerous approaches based on one way hashing and digital signature methods have been introduced so far such as SHA, MD5, etc. [25].

Although in the past researchers preferred data hiding and digital watermarking algorithms, recently new passive approach which does not embed any secondary data into the image has become more popular. The passive (also called blind) methods in contrast to the active ones do not use any prior information about the analyzed image. Numerous algorithms have been proposed in this area, trying to detect image merging [26], traces of non-consistencies in color filter array interpolation [27], traces of geometric transformations, [28], cloning [29], traces of uses of computer graphics [30], JPEG compression inconsistencies [31], etc. All these methods are typically using the fact that digital image editing introduces specific detectable statistical changes into the analyzed image.

The proposed algorithms aim at the minimization of necessary computational power and human intervention. They apply passive

approach and are based on detection of traces of double compression and interpolation, and inconsistencies in chromatic aberrations and in omnipresent noise. We have included as well our method for detection of copy–paste forgeries, when some parts of an image are copied and pasted to other image parts to intentionally hide original image content.

##### 4.1. Detecting double compression

When an image is intentionally altered, it is typically loaded into a photo-editing software and after manipulations are carried out the image is once again re-saved. If the original image was already compressed, during the loading procedure the data are uncompressed and at the re-saving step they are compressed again, potentially with other parameter setting of the compression procedure.

For JPEG compression method this double action can be analyzed and used as an indication of an image alternation, since this re-saving introduces specific changes into the altered image due to the difference of used quantizations matrices of unaltered (primary) and modified (secondary) images.

To understand the core of the proposed algorithm we need to show basics of the JPEG algorithm [32]. The image to be compressed is converted from RGB to YCbCr color channels (Y – luminance, Cb and Cr – chrominance components). Each channel is then split into adjacent blocks of  $8 \times 8$  pixels, transformed by discrete cosine transform (DCT), and quantized. At the end all quantized coefficients are compressed by some variant of Huffman entropy encoding.

The key step is the quantization. Here the  $8 \times 8$  quantization matrix  $Q(u, v)$ , which defines the quantization steps for each used DCT frequency, is applied in lossy manner. The quantized DCT coefficients  $F^Q(u, v)$  are defined as:

$$F^Q(u, v) = \text{round}\left(\frac{F(u, v)}{Q(u, v)}\right), \quad u, v \in \{0, \dots, 7\}, \quad (7)$$

where  $F(u, v)$  are computed DCT coefficients from particular block of individual color channels.

During the double JPEG action the compressions were realized with different quantization matrices  $Q_\alpha$  (primary) and  $Q_\beta$  (secondary). The DCT quantized coefficient is said to be double quantized if  $Q_\alpha(u, v) \neq Q_\beta(u, v)$ . The double quantization is given by:

$$F^{Q^\beta}(u, v) = \text{round}\left(\frac{F^{Q^\alpha}(u, v)Q^\alpha(u, v)}{Q^\beta(u, v)}\right). \quad (8)$$

To determine the presence of double compression artifacts, the method uses a threshold-based quantitative measure. Histograms

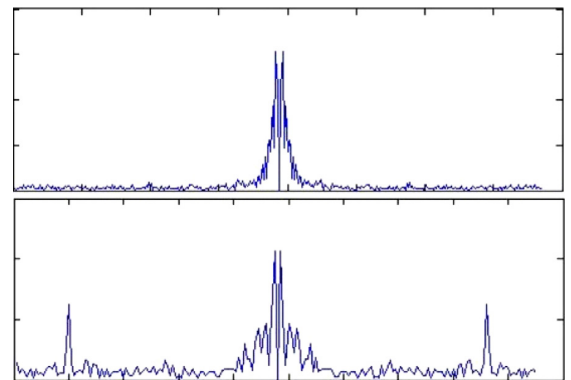


Fig. 5. The top image shows spectrum of a single JPEG image. In the bottom there is shown the spectrum of a double JPEG image.

of DCT coefficients are analyzed in the frequency spectrum (FFT). If their spectral representation exhibits characteristics peak, the image is classified as possibly double JPEG compressed image. The bottom image in Fig 5 demonstrates a typical FFT spectrum of double JPEG images. The image on the top is the spectrum of a single JPEG compressed image. Fig. 15(c) shows an application of the method.

Patterns introduced by double JPEG compression depend on particular compression quality parameters [33]. Proposed method extends the approach introduced in [33], by also allowing to detect double compression locally and thus detecting if two or more images were spliced together. The analyzed image is divided into patches and characteristic peaks are detected patch-wise. The sensitivity to the peak height is adjustable. By varying the sensitivity we can analyze the image and determine if double compression appears non-homogeneously.

It is important to note that detecting the traces of double compression does not necessarily imply the existence of malicious modifications in the image. Often images are re-compressed to achieve smaller size or only simple image adjustment operations such as contrast enhancing were applied. Nonetheless, detecting these changes plays a valuable role in identifying image forgeries.

#### 4.2. Detection of interpolation

When two or more images are spliced together often geometric transformations such as scaling, rotation or skewing of these images are needed in order to create high quality and consistent forgeries. Geometric transformations typically require resampling and interpolation steps. Therefore by having proposed detectors for resampling/interpolation detectors altered images containing tampered portions can be easily identified.

There are two principal steps in geometric transformations. In the first step a spatial rearrangement of pixels of the image is done according to an appropriate transformation function,  $T$ , which maps the coordinates of the input image pixels to points in the output image:

$$x' = T_x(x, y), \quad y' = T_y(x, y). \quad (9)$$

The second step of the geometric transformation deals with an interpolation – resulting intensity values for individual pixel positions in the transformed image are assigned by means of a constructed low-pass interpolation filter  $w$  and rearranged pixels from the input image. Here, to compute signal values at arbitrary locations, discrete samples of the  $f_k$  are multiplied with proper filter weights when convolving them with  $w$ . We denote the result of interpolation operation by  $f^w(x)$ , respectively by  $\mathcal{D}^n\{f^w\}(x)$  which would denote  $n$ th derivative of  $f(x)$ .

By assuming that  $\vartheta$  is an integer, it can be shown that

$$\text{var}\{\mathcal{D}^n\{f^w\}(x)\} = \text{var}\{\mathcal{D}^n\{f^w\}(x + \vartheta\Delta_x)\}, \quad \vartheta \in \mathbb{Z}. \quad (10)$$

In other words it can be shown that interpolation brings into the signal and their derivatives a specific periodicity [28]. This periodicity is dependent on the applied interpolation kernel and can be used for detecting the traces of an interpolation. To this end, a derivative filter is applied to the investigated region,  $b(x, y)$ ,  $\mathcal{D}^n\{b(x, y)\}$  and a Radon transformation is employed in order to find traces of an affine transformation. The Radon transformation computes projections of magnitudes of  $\mathcal{D}^n\{b(x, y)\}$  along specified directions determined by angle  $\theta$  (the projection is a line integral in the given direction), resulting in 180 vectors  $\rho_\theta$ . If the investigated region has been resampled, corresponding auto-covariance sequences defined as  $R_{\rho_\theta}(k) = \sum_i (\rho_\theta(i+k) - \bar{\rho}_\theta)(\rho_\theta(i) - \bar{\rho}_\theta)$  will contain a specific strong periodicity (see Fig. 6 bottom-right). In contrast to [28], only the detection of scale and rotation is

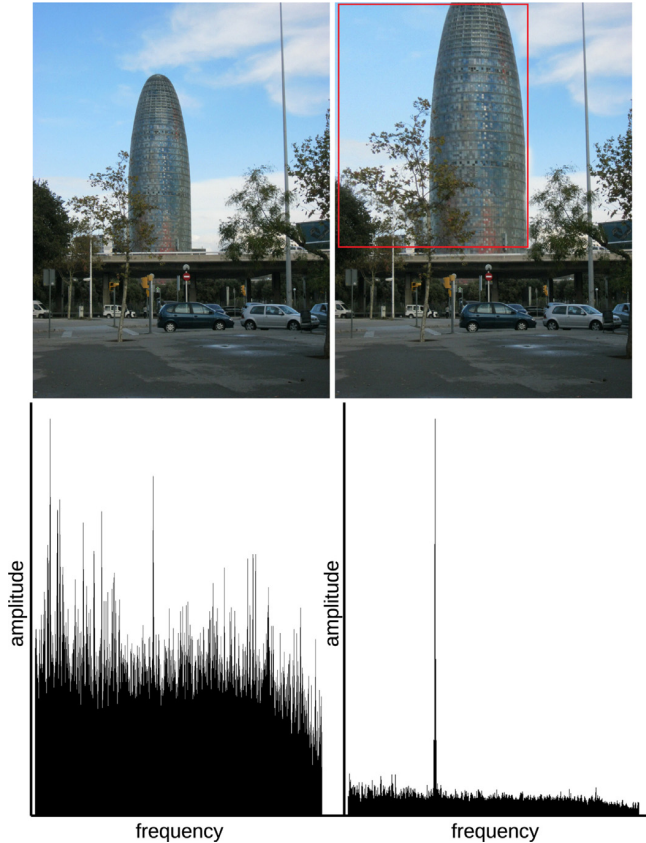


Fig. 6. Detection of interpolation: top-left: the original image stored in the TIFF format; top-right: the image with a re-sized area, denoted by a red box; bottom-left: the result of the method applied on the original image; bottom-right: the result of method applied on the modified image, with a distinctive peak.

supported and results are visualized to the forensic investigator in a single plot so he/she has a full control on decision making.

#### 4.3. Detection of copy-move

One of common types of digital image forgeries is a copy-move forgery [29]. Here, a part of the image is copied, often blurred on its border and pasted into another part of the same image, with the intention to hide an object or a region of the image. Typically, ideal regions for using copy-move forgery are textured areas with irregular patterns, such as grass. Because the copied areas will likely blend with the background it is very difficult for the human eye to detect any suspicious artifacts. Another fact which complicates the detection of this type of tampering is that the copied regions come from the same image, therefore they have similar properties, such as the noise component or color palette. It makes the use of statistical measures to find irregularities in different parts of the image impossible.

Existing copy-move forgery detection methods are mostly based on matching of overlapping image blocks. For example, Fridrich et al. [34] have proposed a method which is based on matching the quantized lexicographically sorted discrete cosine transform (DCT) coefficients of overlapping image blocks. The lexicographical sorting of DCT coefficients is carried out mainly to reduce the computational complexity of the matching step.

The proposed method begins with tiling the image by blocks of  $R \times R$  pixels. Blocks are assumed to be smaller than the size of the duplicated regions, which have to be detected. Blocks are horizontally slid by one pixel rightwards starting with the upper

left corner and ending with the bottom right corner. The next step is about the representation of overlapping blocks with a set of features. The method enables to choose between two different kinds of representation of blocks – either a DCT-based representation of blocks or our representation based on blur moment invariants. The DCT representation of blocks is faster and has a lighter requirement on RAM in comparison to moment invariants, while the blur invariants based method has an advantage in its robustness in scenarios where the copied areas have been intentionally blurred. They are functions of central moments [35,36]. Here, the two-dimensional  $(p+q)$ th order central moment  $\mu_{pq}$  of  $f(x, y)$  is defined as

$$\text{var}\{\mathcal{D}^n\{f^w\}(x)\} = \text{var}\{\mathcal{D}^n\{f^w\}(x + \vartheta\Delta_x)\}, \quad \vartheta \in \mathbb{Z}, \quad (11)$$

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-x_t)^p (y-y_t)^q f(x, y) dx dy, \quad (12)$$

where the coordinates  $(x_t, y_t)$  given by the relations  $x_t = m_{10}/m_{00}$ ,  $y_t = m_{01}/m_{00}$  denote the centroid or the center of gravity of  $f(x, y)$ . Blur invariants are then defined using the recursive relation:

$$B(p, q) = \mu_{pq} - \alpha \cdot \mu_{pq} - \frac{1}{\mu_{00}} \sum_{n=0}^K \sum_{i=m_1}^{m_2} \binom{p}{t-2i} \binom{q}{2i} \cdot B(p-t+2i, q-2i) \mu_{t-2i, 2i}, \quad (13)$$

where

$$\begin{aligned} K &= [(p+q-4)/2], \quad t = 2(K-n+1), \\ m_1 &= \max(0, [(t-p+1)/2]), \\ m_2 &= \min(t/2, [q/2]), \\ \alpha &= 1 \Leftrightarrow p \wedge q \text{ are even}, \quad \alpha = 0 \Leftrightarrow p \vee q \text{ are odd.} \end{aligned}$$

The proposed algorithm uses 24 blur invariants up to the seventh order to create the feature vector  $B = \{B_1, B_2, B_3, \dots, B_{23}, B_{24}\}$  of each block. Using the principal component transformation (preserving the Euclidean distance among blocks) we reduce this dimension.

After the representation of blocks by either DCT or moment invariants, similar blocks are identified. The main assumption here is that a duplicated region consists of many neighboring duplicated blocks. If we find two similar blocks in the analyzed space and if their neighborhoods are also similar to each other, there is high probability that they are duplicated and thus they will be labeled. The output of the algorithm is a map with the same size as the input image (see Fig. 7 bottom), with values either zero (the block at this position is not duplicated) or one (the block at this position is duplicated). The implemented method is based on the optimized analysis of block similarity introduced in [29] resulting in comparable results but faster response.

#### 4.4. Detection of noise inconsistency

Commonly used tool to conceal traces of tampering is the addition of locally random noise to the forged image regions [37]. The amount of noise in an authentic image is usually uniform across the entire image. Adding locally random noise may cause inconsistencies in the image's noise. Therefore, detection of variations of noise levels in an image may signify tampering. Our method is capable of dividing (segmenting) an investigated image into segments with different estimated noise levels. It is based on wavelet analysis followed by tiling high-frequency sub-band into non-overlapping blocks for which noise variation is individually estimated. In the first step a one-level wavelet decomposition [38] of the investigated image is carried out. The high frequency sub-band,  $HH_1$ , gives diagonal details of the image at the highest resolution. Our method tiles this sub-band by non-overlapping blocks  $B_i$  of  $R \times R$  pixels. Blocks are assumed to be much smaller than the size of the corrupted regions to be detected. The size of



**Fig. 7.** Detection of copy-move: Top: the original photo; middle: the tampered image with pasted areas; bottom: the method output based on blur moment invariants showing detected duplicated areas.

blocks can be interactively adjusted. For each block we estimate the standard deviation of the noise. It has been shown [39] that it can be robustly estimated from  $HH_1$  using the median based estimator

$$\hat{\sigma} = \text{median}(|HH_1|)/0.6745.$$

The median measurement is insensitive to isolated outliers of potentially high amplitudes. Often  $\text{median}(|HH_1|)$  is denoted as

MAD( $HH_1$ ) where MAD stands for median absolute deviation. This estimator is very popular and generally provides robust and precise outcomes. Estimated noise levels are visualized (see Fig. 14 right column) in gray level map for identification of potentially tampered areas. The method can be extended by clustering of blocks with similar estimated noise levels [37].

It has to be noted that the noise degradation is the main cause of failure of most existing blind forgery detection methods. These methods are able to work correctly only when the amount of present noise is small. For example, in copy–move forgery additive noise causes mismatches of duplicated regions. This significantly decreases the performance of copy–move forgery detection methods. The same effect can be observed in the methods based on the detection of resampling. Here the noise degradation causes the loss of detectable interpolation based correlation among neighboring pixels.

#### 4.5. Chromatic aberration inconsistency

Camera optical system introduces several kinds of aberration (imperfections) into the image. One of them, chromatic aberration, is caused by failure of a lens to focus all colors to the same convergence point, because lenses have different refractive indices for different wavelengths of light. Although many modern cameras attempt to reduce the effect of chromatic aberration, it is still often manifested as a regular pattern of slight blur and/or color shifts. Deviations from this regularity in a digital photograph can be seen as an indication of tampering.

There are two types of chromatic aberration: longitudinal and lateral. Longitudinal aberration occurs because different wavelengths have different focal planes, therefore all colors cannot be focused (sharp) at once. Lateral aberration occurs because different wavelengths from the same scene point reach the sensor at different positions, therefore different color channels are shifted with respect to each other. The proposed method is based purely on lateral aberration.

It can be derived (e.g., [40]) that the relation between image points of two different color channels (e.g., red channel  $\mathbf{x}_r$  and green channel  $\mathbf{x}_g$ ) corresponding to the same scene point is a simple affine transform with respect to the optical center  $\mathbf{x}_0$  and some isotropic scale  $\alpha$

$$\mathbf{x}_r = \alpha(\mathbf{x}_g - \mathbf{x}_0) + \mathbf{x}_0. \quad (14)$$

The general principle of forgery detection using chromatic aberration is based on the estimation of the unknown parameters  $\{\mathbf{x}_0, \alpha\}$ . They are estimated both globally, using the entire image, and locally, using only a small patch of the image. If the two estimates are not sufficiently similar, the patch is marked as possibly having been tampered with.

Authors of [40] used mutual information as the similarity measure between color channels and found the transform parameters  $\{\mathbf{x}_0, \alpha\}$  using an exhaustive search. Every iteration of such search requires interpolating the entire image and the whole procedure is therefore extremely slow. Motivated by our target application, we propose a method which is much faster and more practical.

Because shifts due to aberrations are small ( $\alpha$  in (14) is very close to 1), the affine transform can be locally approximated by constant shift. We therefore set up a grid of regularly spaced positions throughout the entire image and estimate two sets of the shifts – red to green channel and blue to green channel – in these positions. If the estimated shifts locally deviate from the pattern suggested by Eq. (14), it may be a sign of tampering in this particular area of the image. We only display the estimated pattern of shifts in the form of arrows (see Fig. 15(d)). The final decision whether or not the pattern is sufficiently regular is left to the operator.

The key step of the proposed method is estimation of the shift between two color channels for a small patch around center (grid) pixel. Because the effect of chromatic aberration is relatively small, we must find the shift with high subpixel precision. Naive direct approach would require first upsampling the patch  $N$  times to reach  $1/N$  pixel precision and then performing the registration, which is computationally quite demanding task. The suitable patch sizes depend on a image size, for common 10 megapixel image, our default patch size is  $125 \times 125$  px, resulting in  $29 \times 21$  patches. Processing such image with  $1/100$  subpixel precision would then require performing two interpolations and the registration of  $12,500 \times 12,500$  pixels image (156 megapixel) 609 times for just one of two combinations of color channels, which is computationally prohibitive.

To achieve better performance, we propose modification to the brute force method of [40]. First, we replace mutual information with direct correlation as the similarity measure for registration. Second, we calculate the correlation in the Fourier domain and use the two-step discrete Fourier method (DFT) method described in [41], in which the required upsampling can be done by virtually zero-padding the Fourier transforms of the two patches. The respective shift between two color channels is then estimated by the following procedure:

1. Calculate Fourier transforms of both color channels in the examined patch.
2. Perform correlation in the Fourier domain by pixel-wise multiplication.
3. Perform inverse Fourier transform taking into account the fact that the Fourier transform has been virtually zero-padded to  $N$  times their original size, where  $1/N$  is the required subpixel precision.
4. Find coordinates of maximum, which is the sought shift (when divided by  $N$ ).

For the inverse Fourier transform in the step 3 we do not use FFT, which would require physically zero-padding the patches, but rather express the 2D Fourier transform as a matrix multiplication of the patch in the form

$$p = RPC, \quad (15)$$

where  $p$  and  $P$  are the image patch (one of the color channels) and its Fourier transform, respectively, and  $R$  and  $C$  are the row and column inverse of the 2D DFT matrix. In the formation of the  $R$  and  $C$  matrices we take into account the fact that, firstly, we need the inverse FT only in the limited number of coordinates, expecting the inter-channel shift close to zero, and, secondly, that the FTs of the patches have been zero-padded so we include only non-zero input in the calculation.

Using the described procedure, the shift can be accurately estimated in a fraction of time (and memory) required for the naive approach. To further speed-up the process, we first estimate the shift with  $1/4$  pixel precision and only in its neighborhood do we refine the estimate to the full  $1/100$  precision.

#### 5. Image and video restoration

The main goal of image and video restoration in forensic analysis is to increase the perceivable image content to the end user. This is a frequent situation since videos from surveillance cameras often do not have sufficient resolution and sharpness and are often blurred and noisy. The restoration module is a powerful tool that addresses these issues. It is capable of removing noise and compression artifacts, and if multiple images of the same object (e.g., several frames from a video sequence) are available it also increases spatial resolution. The module consists of three

algorithms that can be run independently or applied sequentially to corrupted input data. Below we describe each method and provide insights into the applied mathematical tools.

### 5.1. Denoising

Denoising is the most common image and video restoration task. In video surveillance two sources of noise impair the quality of acquired images. We refer to them as external and internal. Two examples of the external source are fog and haze. Various methods were proposed to remove such degradation, recently e.g., [42], but they are not considered here. Instead we focus on the internal source which is intrinsic to any acquisition device. The internal noise is a mixture of Gaussian and Poissonian noises, we approximate the noise for the image processing purposes by the additive Gaussian noise. We write this symbolically

$$g = u + n, \quad (16)$$

where  $u$  is the original image,  $g$  the observed noisy image and  $n$  the Gaussian noise with normal distribution  $N(0, \sigma^2)$ .

Our choice of denoising method was influenced by the generally valid trade-off between the quality of restoration and the time needed for computation. One group of the best performing fast methods is based on thresholding of wavelet coefficients [43]. In the implemented application we exploited special properties of the dual-tree complex wavelets DT-CWT described in [44]. This type of wavelets has more isotropic behavior than standard wavelets that emphasize vertical and horizontal edges, and weaken diagonal edges. While slightly more time-consuming than standard orthogonal wavelets, the time of computation is still linear in the number of pixels.

Formally, the estimated image equals

$$u = W^{-1} \arg \min_v \frac{1}{2\sigma^2} \|g - W^{-1}v\|^2 + \alpha \|v\|_1, \quad (17)$$

where  $W$  is the wavelet transform,  $W^{-1}$  its inverse,  $v$  the wavelet coefficients of the solution,  $\sigma^2$  noise variance and  $\alpha > 0$  a parameter controlling the level of smoothness. Thanks to the special property of complex wavelets being the Parseval frame, the solution can be computed in linear time with respect to the number of pixels by soft thresholding

$$S_\alpha(a) = \begin{cases} a - \alpha & a > \alpha \\ 0 & |a| \leq \alpha \\ a + \alpha & a < -\alpha, \end{cases} \quad (18)$$

applied on individual coefficients of the wavelet transform, i.e.,  $u = W^{-1} S_{\alpha\sigma^2}(Wg)$ .

To achieve the optimal smoothness, the estimated noise variance  $\sigma^2$  can be adjusted by user. Fig. 8 demonstrates our result for a very noisy input.

### 5.2. Super-resolution

Super-resolution (SR) algorithm fuses multiple input low-resolution images (video frames) and estimates the latent high-resolution image (frame). We model the acquisition process and then apply an inverse method to recover the latent image as we proposed originally in [45]. The nature of degradation in the acquisition process implies that this is an ill-posed problem and we need additional regularization. Appropriate numerical methods, such as [46], are necessary to solve non-linear equations, which are the outcome of the regularized functional.

The acquisition process models the observed low-resolution image  $g_k$  as the downsampled ( $D$ ) and warped ( $T_k$ ) high-resolution

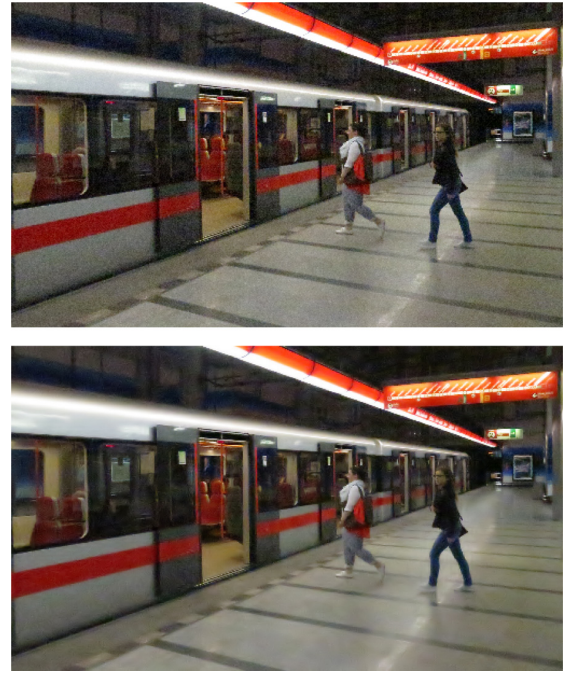


Fig. 8. Denoising: Top: the original image taken in poor lighting conditions causing profound noise; bottom: the output of the proposed restoration.

image  $u$

$$g_k = DT_k u + n, \quad (19)$$

where  $n$  is additive noise. The downsampling operator  $D$  mimics the sampling phenomenon taking place in the camera sensor by performing convolution with a sensor blur and subsampling the data. The operator  $D$  is the same for all the observed images  $g_k$ 's and is fully determined by user parameters, which are the sensor blur size and subsampling factor. In our case, we use gaussian blur as the sensor blur as proposed in [47]. The subsampling factor is equivalent to the SR factor and specifies by what ratio we want to increase the resolution of input images. The warping operator  $T_k$  geometrically transforms  $u$  to be aligned with  $g_k$ . The acquisition model assumes that all the observed images  $g_k$ 's differ only by the geometric transformations  $T_k$ 's, which are however unknown.

The critical step in the SR algorithm is to correctly estimate  $T_k$ 's. Typically we choose one input image  $g_r$  as a reference image, the corresponding  $T_r$  is identity and all other  $T_k$ 's are calculated towards the reference image. For this purpose we use an optical flow (OF) algorithm, which estimates local shifts (motion field) between two images with sub-pixel accuracy. The estimated motion field then fully defines the warping matrix  $T_k$ . There is a vast number of OF algorithms and several benchmarks that compare their performance [48]. For our purpose we have chosen a method in [49], which has the best ratio of precision to time complexity, and included modifications suggested in [50]. The implementation of the OF algorithm is pyramidal, i.e., calculating shifts on multiple scales, which allows accurate estimation of large and small shifts simultaneously. Smoothness of the estimated motion field is forced by the Total Variation (TV) regularization. If the input images contain artifacts, such as noise or compression artifacts, the motion field often contains outliers negatively influencing the SR step. We have improved the OF algorithm by adding an optional constraint for parametric motion field models. The user can choose between two parametric models: translation and affine. The translation model constrains the estimated motion field to give one global translation vector. Such scenario is typically

useful when we are interested in SR of a relatively small region where the geometric transformation among images is well approximated by translation, e.g., a license plate of a moving car. The affine model constrains the field to be represented by a six-parameter linear transform, which is the most general 2D linear transform. Note that this allows for rotation and scale changes in the input images.

Once the warping operators  $T_k$ 's are estimated, we proceed to solve (19). Following the Bayesian paradigm, the optimal solution is the maximum a posteriori (MAP) estimator. Let the noise  $n$  be normally distributed and the image prior be a sparsifying distribution  $\exp\{-\lambda\|Cu\|_p^p\}$ , where  $p \leq 1$  and  $C$  is a filtering operator returning features that are assumed to be sparse. A common example of  $C$  is image gradient. The MAP estimator is then the minimum of the energy function

$$E(u) = \frac{1}{2} \sum_k \|DT_k u - g_k\|^2 + \lambda \|Cu\|_p^p. \quad (20)$$

The first term of  $E$  is called a data term and the second one is regularization. Due to the sparsity measure  $\|Cu\|_p^p$ , derivatives of  $E$  are non-linear and we thus apply a linearization method referred to as the half-quadratic algorithm [51]. The problematic term  $\|Cu\|_p^p$  is replaced with a quadratic form  $\langle u, Lu \rangle$ , where  $L$  must be iteratively updated with previously calculated  $Cu$ . The action of  $L$  can be interpreted as space-variant convolution with a Laplacian-like filter, of which coefficients spatially vary base on  $Cu$ . In addition, the action of  $DT_k$  is also space-variant convolution with a filter (sensor blur) which is shifted by sub-pixel vectors in  $T_k$ . In the half-quadratic algorithm we thus iteratively solve for  $u$  a linear system

$$\left( \sum_k T_k^* D^* D T_k + \lambda L \right) u = \sum_k T_k^* D^* g_k, \quad (21)$$

for which we use a Conjugate Gradient (CG) method. A fast implementation using FFT is not possible in this case since  $\sum_k T_k^* D^* D T_k + \lambda L$  is space-variant convolution. However CG methods are relatively efficient as they typically require a small number of iterations (around 10) for this type of problems.

In Fig. 9 we compare default camera demosaicing with standard SR reconstruction of JPEG images and our SR reconstruction of RAW images. The demosaicing algorithm applies on a single image whereas the SR algorithms take multiple images, therefore SR is expected to provide better results as more information are at disposal to process. The results are cropped to better see and compare the reconstruction quality of the car license plate. The camera demosaicing output in (b) was spatially interpolated by a factor of 2 to obtain the resolution of SR outputs. The demosaicing images are typically used in practice, since they are the default output of any digital camera. The demosaicing process removes color filter array (also called Bayer filter mosaic) and also performs other tasks, such as denoising and color correction, which alters the images in a way not suitable for SR. If we apply standard SR algorithms on such images, we obtain a suboptimal result in (c). In this case the SR factor was 2. However, our SR algorithm is adapted to use directly RAW images and work with Bayer mosaic, which outperforms standard SR as can be seen in (d). The resolution of RAW images is half the resolution of demosaicing images and therefore the SR factor in the second case (d) is 4 to obtain the same resolution as in the first case (c). The license plate is difficult to read after demosaicing and even after applying SR on these images. Our SR on RAW images shows dramatic improvement with all the letters and numbers clearly legible.



(a) RAW input images



(b) Camera demosaicing – JPEG



(c) Super-resolution of JPEG



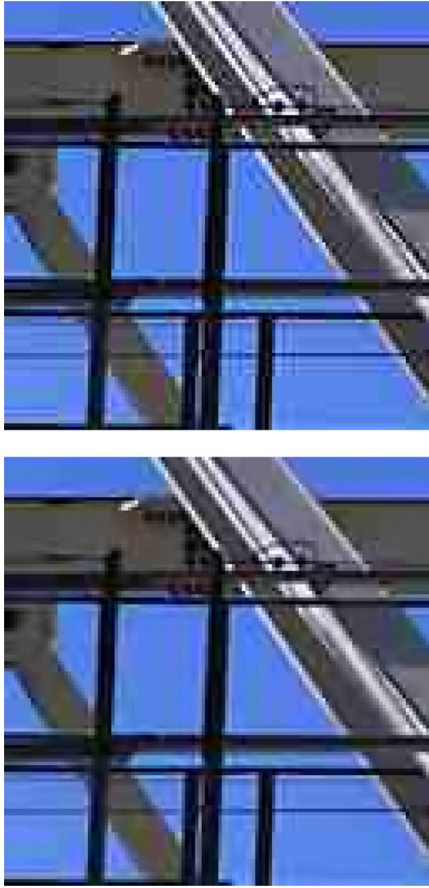
(d) Super-resolution of RAW

**Fig. 9.** Super-resolution: (a) sequence of photos captured by a DSLR camera; (b) camera default demosaicing JPEG output of one photo; (c) super-resolution reconstruction from demosaicing JPEG images; (d) super-resolution reconstruction directly from RAW images.

### 5.3. JPEG artifact removal

Original JPEG standard [52] is based on compression of cosine transform coefficients. Since compression works independently on individual blocks of  $8 \times 8$  pixels, higher compression ratios result in a visually disturbing checkerboard pattern and artifacts along strong edges (see Fig. 10) [53]. Although there is a newer compression standard JPEG 2000, the original version still prevails because of its simplicity and speed. For this reason, our algorithm was designed specifically to work with image compressed by the original standard.

For the removal of JPEG artifacts we proposed to use an approximation of the Bayesian MAP approach, minimizing

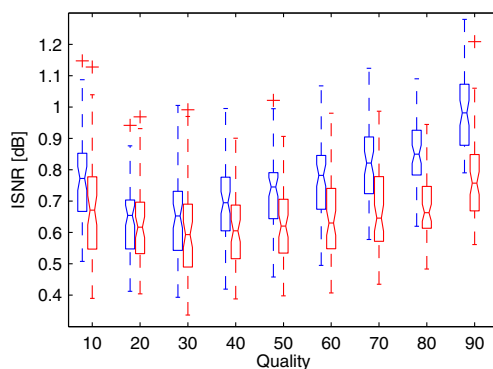


**Fig. 10.** JPEG restoration: top: the image compressed by JPEG with quality factor 20; bottom: the output of our reconstruction.

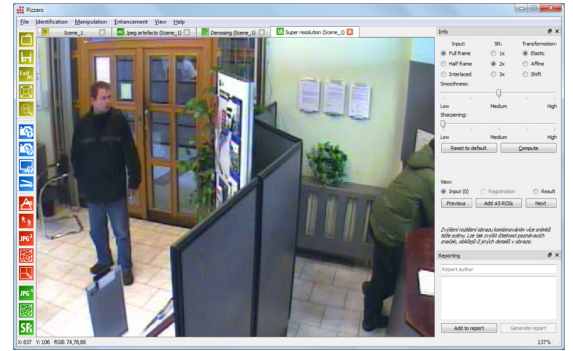
simultaneously the  $l_1$ -norm of wavelet coefficients and discrepancy from the quantization constraint  $QCu \in (QCg - 0.5, QCg + 0.5)$

$$\arg \min_u \frac{1}{2} \|QCg - QCu\|^2 + \alpha \|Wu\|_1, \quad (22)$$

where  $C$  is the orthogonal matrix of block discrete cosine transform and  $Q$  the diagonal matrix containing inverted elements of the



**Fig. 11.** Statistical comparison of our algorithm and [55] on a database of 60 images. All images were compressed into JPEG format and then decompressed/restored by both algorithms. This was done for nine JPEG quality levels, from 10 to 90. For each quality, our results is shown on the left (blue) and [55] on the right (red). The graph shows the median value, first and third quartiles, whiskers extending to the most extreme data points not considered outliers, and outliers plotted individually. The notches indicate the confidence intervals of the median.



**Fig. 12.** An example of the PIZZARO user interface window.

quantization table defined by the JPEG standard [52]. For regularization, we used the same wavelet transform  $W$  as for image denoising [44], giving results superior to standard wavelets or total variation [54,55]. Fig. 11 shows a statistical comparison of the proposed novel method in terms of SNR with one of recent methods [55] based on total variation.

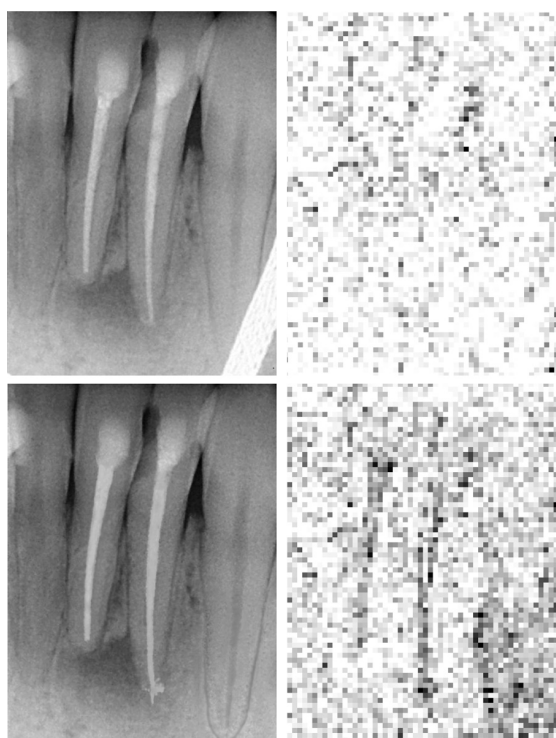
## 6. PIZZARO software

All proposed algorithms have been included to PIZZARO [1], the set of software tools for interactive image and video forensic analysis and an application of individual procedures (see Fig. 12 for the user interface). Software PIZZARO was developed in C++ using Qt toolkit for graphical user interface, OpenCV for image manipulation and several libraries for reading specific image/video formats (libpng, libjpeg, libxif, FFmpeg), and can be run under any version of Windows® from XP up to 8.1 in either 32-bit or 64-bit mode. As no platform specific libraries are used, it is possible to compile the software for other platforms (e.g., OS X, or specific Linux distributions).

It is designed to allow efficient work with images and videos to quickly reveal and apply desired methods, with the possibility to



**Fig. 13.** Gas station super-resolution use case: top: an input video sequence recorded by a security camera with close-ups of suspect's left and right hand; bottom: an estimated high-resolution image with two close-ups. Recovered image suggests that the suspect wore a ring and hold the handgun in an unconventional way using the middle finger on a trigger.



**Fig. 14.** X-ray scans image content verification use case: left column: dental X-rays from both parties to the lawsuit; right column: the output of the noise inconsistency detection evaluated on the X-rays, the bottom image manifests suspicious structures.

open many images or videos at the same time, compare results achieved using different parameters, view Exif metadata information for JPG images, etc.

The results can be immediately exported to HTML by the integrated reporting system for printing. This functionality enables to log and document how were the results achieved together with applied parameters, operator identity, file paths and dates and other possible settings. All this assures the repeatability of the applied procedures.

## 7. Use cases

The proposed software system has been successfully applied on several real cases of police investigation in cooperation with different police departments in the Czech Republic.

An example of the utilization of the image and video restoration module is the *Gas station* case. In Fig. 13 there is an output of the video restoration that assisted in a police investigation conducted by the Czech Police Department in Liberec, Division of Public Criminality, which resulted in convicting a suspect of burglary and homicide. Cropped frames from a video sequence recorded by a security camera are on left. The close-ups (top and bottom) of the left and right hand of the suspect show insufficient resolution of the video sequence, where details are missing. They were significantly restored after processing the video with proposed PIZZARO video restoration module (right), which increased the resolution and diminished the omnipresent noise. The most notable is the presence of a golden ring on suspect's left hand and an unconventional way of holding a handgun with a protruding index finger, which suggested that the suspect used his middle finger on a trigger.

Another example of the application of the proposed procedures is related to the demonstration that the photograph was tampered



(a)



(b)



(c)



(d)

**Fig. 15.** Billboard image content verification use case: (a) the original photo; (b) the image with a modified area; (c) the output of the double-compression detection; (d) the output of the chromatic aberration inconsistency detection.

in order to create false evidence. The analyzed image data were two dental X-ray scans capturing the same patient (Fig. 14, left column). They were supposed to be used to validate applied medical procedure and they were used by opposite parties to the

dispute. PIZZARO methods were here applied to investigate the originality of the data. We evaluated the noise consistency results of which are demonstrated in Fig. 14 in the right column. While the noise structure of the upper X-ray scan does not relate to the image content, the bottom X-ray noise is clearly related to the image content, which is very unlikely. The conclusion that the bottom X-ray is fraud was then even more significantly confirmed by the fact, that these two scans are geometrically identical apart from the areas where the noise inconsistencies were manifested (the procedure of the X-ray data acquisition makes it almost absolutely impossible to obtain two spatially identical scans).

**Billboard**, the last example of the PIZZARO application, comes from the area of investigation of tax evasion. Here, the question to be answered is whether the photographs of billboards justifying the tax payer costs on his marketing campaign are genuine, capturing real billboards (see Fig. 15(a)) and are not just results of Photoshop work when the fake billboard image content was copied into the real scenery (Fig. 15(b)). The latter case represents tax fraud, because the tax payer is claiming costs which were not spent in reality. Fig. 15(c) demonstrates the output of the double compression detection module, where the dark area marks parts of the photograph with suspicious parameters settings. The same area was identified by the procedure testing the consistency of chromatic aberrations (see Fig. 15(d), the arrows have random direction and size).

The experience with cases analyzed by PIZZARO software has shown that it is difficult to provide some defined manual how to proceed with various tasks. The variability of possible scenarios is large. Moreover, the three modules of the proposed PIZZARO toolbox are each oriented on different problems, so not too often methods from more modules are used on the same dataset. There are some limitations which should be taken into account not to decrease the efficiency of the methods. Denoising or JPEG artifact removal should not be applied before any of the other method, otherwise mathematical assumptions of the algorithms will be violated. Denoising can be achieved by superresolution (SR) too, so the denoising of the output of SR does not make sense. Even more difficult situation arises when the forensic case under investigation is to question the statement that the photograph was tampered. Methods verifying the authenticity of image content often provide only an indication that some parts of the photographs were tampered because they manifest suspicious characteristics. The recommended practice is to apply several methods for content verification in order to get stronger evidence.

## 8. Conclusion

We have introduced a set of methods for image and video forensic analysis. This work was based on tight cooperation of scientists from the Institute of Criminalistics, National Drug Headquarters of the Criminal Police and Investigation Service of the Police of the Czech Republic, and image processing experts from the Czech Academy of Sciences and thus reflecting best practices used in the criminal investigation utilizing images and/or videos. We have addressed two key problems identified as the most important issues of which automation can facilitate criminalists work. They are verification of image/video data with respect to their credibility and origin and image restoration aimed at diminishing unwanted artifacts and at the same time increasing the data quality using all available information.

The described methods include the novel approaches for: an assignment of the images to the source device, the image content verification methods based on various image characteristics, and finally, methodology for removing compression artifacts and even increasing the data resolution. The functionality of some methods was illustrated by means of the comparison to other existing

approaches and by demonstrating performance tests. Next to this, the detection of LCD image re-capture and noise removal were implemented in order to provide required functionality. Proposed methods were designed or modified in order to fulfill time, quality, and interactivity criteria. It should be noted that the output of the proposed methods has often an indicative nature thus the recommended practice is to apply several methods to get stronger evidence.

All theoretical results were implemented in the PIZZARO software tool (available for testing and licensing), which consists of the image processing functionality and reporting and archiving functions to ensure the repeatability of image analysis procedures, fulfilling formal aspects of the analysis work.

## Acknowledgements

The work was supported by the Czech Ministry of the Interior, Grant No. VG20102013064 and by GACR Grant No. GA13-29225S.

## References

- [1] PIZZARO – Tools for Imaging Device Identification, Authentication, and Image Reconstruction, <http://pizzaro.utia.cas.cz>.
- [2] Access Data, <http://accessdata.com/solutions/digital-forensics/forensic-toolkit-ftk>.
- [3] Fourandsix, <http://www.fourandsix.com/>.
- [4] Verifeyed, <http://www.verifeyed.com/>.
- [5] Belkasoft, <http://belkasoft.com/en/>.
- [6] Forensic Pathways, <http://www.forensic-pathways.com/products-and-services/forensic-image-analyser>.
- [7] Error Level Analysis, <http://www.errorlevelanalysis.com/>.
- [8] Amped, <http://ampedsoftware.com/>.
- [9] M. Jerian, S. Paolino, F. Cervelli, S. Carrato, A. Mattei, L. Garofano, A forensic image processing environment for investigation of surveillance video, *Forensic Sci. Int.* 167 (2–3) (2007) 207–212.
- [10] Motion DSP, <http://www.motionsdp.com/products/lkenalSR>.
- [11] Cognitech, <https://cognitech.com/>.
- [12] RTC Vision, <http://www.rtc-vision.com/>.
- [13] Impress, <http://www.impress4video.com/>.
- [14] Reproducible Research, <http://reproducibleresearch.net/super-resolution>.
- [15] J. Lukas, J. Fridrich, M. Goljan, Digital camera identification from sensor pattern noise, *IEEE Trans. Inf. Forensic Secur.* 1 (2) (2006) 205–214.
- [16] M. Chen, M. Goljan, J. Lukas, Determining image origin and integrity using sensor noise, *IEEE Trans. Inf. Forensic Secur.* 3 (1) (2008) 74–90.
- [17] M. Chen, J. Fridrich, M. Goljan, J. Lukas, Source digital camcorder identification using sensor photo response non-uniformity, in: *Proc. SPIE*, 6505, 2007, pp. 65051G–65051G-12.
- [18] B. Mahdian, A. Novozamsky, S. Saic, Determination of stop-criterion for incremental methods constructing camera sensor fingerprint, in: *13th Int. Workshop on Digital-Forensics and Watermarking*, Taipei, Taiwan, 2014.
- [19] H. Cao, A.C. Kot, Identification of recaptured photographs on LCD screens, in: *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, IEEE, 2010, 1790–1793.
- [20] T. Ojala, M. Pietikinen, T. Menp, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (7) (2002) 971–987.
- [21] M. Kharrazi, H. Sencar, N. Memon, Blind source camera identification, in: *2004 International Conference on Image Processing, ICIP'04*, vol. 1, 2004, 709–712.
- [22] Y. Chen, Z. Li, M. Li, W.-Y. Ma, Automatic classification of photographs and graphics, in: *2006 IEEE International Conference on Multimedia and Expo*, 2006, 973–976.
- [23] B. Mahdian, R. Nedbal, S. Saic, Blind verification of digital image originality: a statistical approach, *IEEE Trans. Inf. Forensic Secur.* 8 (9) (2013) 1531–1540.
- [24] M. Arnold, M. Schmucker, S.D. Wolthusen, *Techniques and Applications of Digital Watermarking and Content Protection*, Artech House, Inc., Norwood, MA, USA, 2003.
- [25] J. Katz, Y. Lindell, *Introduction to Modern Cryptography* (Chapman & Hall/CRC Cryptography and Network Security Series), Chapman & Hall/CRC, 2007.
- [26] Z. Lint, R. Wang, X. Tang, H.-Y. Shum, Detecting doctored images using camera response normality and consistency, in: *CVPR'05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) – vol. 1*, IEEE Computer Society, Washington, DC, USA, 2005, pp. 1087–1092.
- [27] A. Popescu, H. Farid, Exposing digital forgeries in color filter array interpolated images, *IEEE Trans. Signal Process.* 53 (10) (2005) 3948–3959.
- [28] B. Mahdian, S. Saic, Blind authentication using periodic properties of interpolation, *IEEE Trans. Inf. Forensic Secur.* 3 (3) (2008) 529–538.
- [29] B. Mahdian, S. Saic, Detection of copy-move forgery using a method based on blur moment invariants, *Forensic Sci. Int.* 171 (2–3) (2007) 180–189.

- [30] A.E. Dirik, S. Bayram, H.T. Sencar, N. Memon, New features to identify computer generated images, in: IEEE International Conference on Image Processing, ICIP'07, vol. 4, 2007, 433–436.
- [31] J. Fridrich, T. Pevny, Detection of double-compression for applications in steganography, IEEE Trans. Inf. Secur. Forensics 3 (2) (2008) 247–258.
- [32] W.B. Pennebaker, J.L. Mitchell, JPEG Still Image Data Compression Standard, Kluwer Academic Publishers, Norwell, MA, USA, 1992.
- [33] B. Mahdian, S. Saic, Detecting double compressed JPEG images, in: The 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP-09), London, UK, 2009.
- [34] J. Fridrich, D. Soukal, J. Lukas, Detection of copy-move forgery in digital images, in: Proceedings of Digital Forensic Research Workshop, IEEE Computer Society, Cleveland, OH, USA, 2003, pp. 55–61.
- [35] J. Flusser, T. Suk, Degraded image analysis: an invariant approach, IEEE Trans. Pattern Anal. Mach. Intell. 20 (6) (1998) 590–603.
- [36] J. Flusser, T. Suk, S. Saic, Image features invariant with respect to blur, Pattern Recognit. 28 (11) (1995) 1723–1732.
- [37] B. Mahdian, S. Saic, Using noise inconsistencies for blind image forensics, Image Vis. Comput. 27 (10) (2009) 1497–1503.
- [38] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (7) (1989) 674–693.
- [39] D. Donoho, I. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 8 (1994) 425–455.
- [40] M.K. Johnson, H. Farid, Exposing digital forgeries through chromatic aberration, in: Proceedings of the 8th Workshop on Multimedia and security, ACM, 2006, pp. 48–55.
- [41] M. Guizar-Sicairos, S.T. Thurman, J.R. Fienup, Efficient subpixel image registration algorithms, Opt. Lett. 33 (2) (2008) 156–158.
- [42] K.B. Gibson, T.Q. Nguyen, An analysis and method for contrast enhancement turbulence mitigation, IEEE Trans. Image Process. 23 (7) (2014) 3179–3190, <http://dx.doi.org/10.1109/TIP.2014.2328180>.
- [43] S. Mallat, A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way, 3rd ed., Academic Press, 2008.
- [44] N. Kingsbury, Complex wavelets for shift invariant analysis and filtering of signals, Appl. Comput. Harmon. Anal. 10 (3) (2001) 234–253.
- [45] F. Sroubek, G. Cristobal, J. Flusser, A unified approach to superresolution and multichannel blind deconvolution, IEEE Trans. Image Process. 16 (2007) 2322–2332.
- [46] J. Eckstein, D.P. Bertsekas, On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators, Math. Program. 55 (1992) 293–318.
- [47] D. Capel, Image Mosaicing and Super-Resolution (Cphc/Bcs Distinguished Dissertations), SpringerVerlag, 2004.
- [48] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, R. Szeliski, A database and evaluation methodology for optical flow, Int. J. Comput. Vis. 92 (1) (2011) 1–31.
- [49] C. Liu, Beyond pixels: Exploring new representations and applications for motion analysis (Ph.D. thesis), Massachusetts Institute of Technology, 2009.
- [50] D. Sun, S. Roth, M. Black, Secrets of optical flow estimation and their principles, in: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, 2432–2439.
- [51] G. Aubert, P. Kornprobst, Mathematical Problems in Image Processing, Springer Verlag, New York, 2002.
- [52] JPEG File Interchange Format (JFIF), Tech. rep., ECMA TR/98, Ecma International, 2009.
- [53] B. Gunturk, Y. Altunbasak, R. Mersereau, Multiframe resolution-enhancement methods for compressed video, IEEE Signal Process. Lett. 9 (6) (2002) 170–174.
- [54] F. Alter, S. Durand, J. Froment, Adapted total variation for artifact free decomposition of JPEG images, J. Math. Imaging Vis. 23 (2005) 199–211.
- [55] K. Bredies, M. Holler, A total variation-based jpeg decompression model., SIAM J. Imaging Sci. 5 (1) (2012) 366–393.

# Detecting Cyclostationarity in Re-Captured LCD Screens

Babak Mahdian\*, Adam Novozamsky and Stanislav Saic

*Institute of Information, Theory and Automation of the Czech Academy of Sciences, Prague, Czech Republic*

## Abstract

It is easy to display a video or picture on an LCD screen and recapture it by using a camera to hide traces of digital image manipulation or fool an access system based on face recognition technique. In this paper, we show that humans do not have a good performance in detecting recaptured data from LCD screens. Hence, it is important to have methods capable of distinguishing between natural videos and pictures and those recaptured ones. In this paper, we show that, typically, recaptured images and videos from LCD screens exhibit detectable periodic patterns that are caused by regular sampling grid of the LCD monitor and aliasing. We develop our method using the theory of cyclostationarity and experimentally validate it. The term cyclostationarity refers to a special class of signals which exhibit periodicity in their statistics. Our method will be based on the fact that a cyclostationary signal has a frequency spectrum correlated with a shifted version of itself.

**Keywords:** Biometrics; Aliasing; Cyclostationarity; CFA; Image and video recapturing; Digital forensics

## Introduction

In recent years, we have observed a dynamic development in two essential areas of forensics analysis of digital images and videos integrity verification (genuineness analysis) and image ballistics (source device verification). Although past research in the areas of image integrity verification and image ballistics have mainly focused on data hiding and digital watermarking approaches [1,2], today there is a growing alternative approach called the passive one which does not need embedding any secondary data into the image [3]. Methods published in this area have, for example, attempted to detect image splicing [4-6], traces of inconsistencies in color filter array interpolation [7-9], traces of geometric transformations [10-12], cloning [13-15], computer graphics generated photos [16-18], JPEG compression inconsistencies [19-21], file structure inconsistencies [22,23], etc. Typically, pointed out methods are based on the fact that digital image editing brings specific detectable inconsistencies and statistical changes into the image.

In this paper we will deal with automatic recognition of videos and pictures that have been recaptured from LCD monitors. Recent advances in digital camera technology have meant that high resolution images can easily be obtained at a relatively low cost by using digital camera and smart phones. Moreover, a widespread availability of high quality softcopy display mediums, such as LCD monitors have made it possible to reproduce digital images with ease by recapturing the photo or the video from a display using a digital camera.

The motivation of detecting re-captured images and videos can be several. For example, automatic distribution of illegally captures movies from LCD screens. Recapturing is an easy tool to eliminate copy-right related invisible watermarks hidden in images and videos. Another area that needs to be capable of detecting recaptured videos and images is the authentication area. Access systems using face recognition techniques are often vulnerable to spoofing attacks. In a spoofing attempt, a person tries to masquerade as another person to gain an access to the system.

Our motivation does come from the digital forensics point of view. A large portion of digital forensics methods are bases on searching for inconsistencies among pixel. This, they can easily be overcome by recapturing a digitally manipulated image or video. The re-captured image would not contain traces of digital manipulation

and inconsistencies among pixels. In other words, it would act as an original image. In other words, the forger can display fake images on LCD display and recapture the manipulated digital image to overcome image forensic systems. Consequently, detecting recapturing can signify tampering.

This paper will introduce a method capable of detecting recaptured images and videos using a single image and a single frame analysis. Hence, from now on, we will only consider digital images. In case of videos, the method can easily be applied on individual frames separately.

We also will show the results of an experiment showing human performance in identification of LCD re-captured images. We will see that recaptured images from the LCD screen are often perceptually indistinguishable to humans. However, there are fine differences between LCD screen recaptured images and non-recaptured ones caused, for example, by regular monitor pixel grid projected into the recaptured image and aliasing. We will use these differences to develop the method. Specifically, we will detect periodic properties present in the LCD recaptured images by using theory and methods of cyclostationarity. The term cyclostationarity refers to a special class of signals which exhibit periodicity in their statistics. Our methods will be based on the fact that a cyclostationary signal has a frequency spectrum correlated with a shifted version of itself.

The rest of the paper is organized as follows. In section 2 we will show related work. In section 3 we will perform an experiment to better understand the human performance in identification of LCD re-captured images. In section 4 we show differences between LCD screen recaptured images and non-recaptured ones and introduce a method to detect them. In section 5 we run an experiment to validate the method. In the last section, we summarize this paper by discussing the results and pointing out conclusions.

**\*Corresponding author:** Babak Mahdian, Institute of Information, Theory and Automation of the Czech Academy of Sciences, Pod Vodárenskou věží 4, CZ-182 08, Prague 8, Czech Republic, Tel: +420 266 053 111, E-mail: [mahdian@utia.cas.cz](mailto:mahdian@utia.cas.cz)

**Received** March 30, 2015; **Accepted** July 29, 2015; **Published** July 31, 2015

**Citation:** Mahdian B, Novozamsky A, Saic S (2015) Detecting Cyclostationarity in Re-Captured LCD Screens. J Forensic Res 6: 294. doi: [10.4172/21577145.1000294](https://doi.org/10.4172/21577145.1000294)

**Copyright:** © 2015 Mahdian B, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## Related Work

The problem of detecting and identifying recaptured images from LCD monitors has received interest from researchers in recent years and we have been observing a growing number of publications in this area.

Ke Yongzhen et al. have used combinations of low-level features including texture, noise, difference histogram, and color information to train a support vector machine classifier capable of detecting recaptured images. Hani Muammar and Pier Luigi Dragotti [24] have searched for the presence of aliasing due to sampling of the monitor pixel grid to identify recaptured images. To validate their approach, an investigation into the aliasing introduced in a digitally recaptured image has been conducted. Xinting Gao et al. [25] have introduced a recaptured image database captured by smart phone cameras. Huacheng Liu and Rangding Wang have proposed to identify recaptured images using DCT coefficients. They have noticed that the low frequency of an image mainly reflects such information as texture and profile details. So, they described the difference between real images and recaptured images by using the low frequency of DCT coefficients. Thirapiroon Thongkamwitoon et al. [26] have used the fact that edge profiles of single and recaptured images are different in order to detect recaptured images. They have trained two alternative dictionaries using the K-SVD approach to achieve their goal. Neslihan Kose and Jean-Luc Dugelay [27] have proposed an anti-spoofing approach, which is based on analysis of contrast and texture characteristics of captured and recaptured images. A method based on a rotation invariant local binary pattern variance has been used. Xinting Gao et al. [28] have presented a physical model for image recapturing. Their motivation was to make robot vision more intelligent and make a single-image-based countermeasure for re-broadcast attack on a face authentication system feasible. M. Visentini-Scarzanella, Dragotti PL [29] has derived a curve model for straight lines deformed after single capture under a radial distortion model and recapturing. Bestagini P et al. [30] have proposed a detector based on the analysis of a characteristic ghosting artifact left by the recapture process. Xiaoyang Tan et al. have used the Lambertian model to propose strategies to extract the information about different surface properties of a live human face or a photograph. Jiangwei Li et al. [31] have used structure and movement information of live face to propose a live face detection based on the analysis of Fourier spectra. Hong Cao and Kot AC [32] have proposed to use a set of statistical features to capture the common anomalies introduced in the camera recapturing process on LCD screens. Jiamin Bai et al. [33] have focused on paper recapturing and used micro-textures present in printed paper to detect images recaptured from printed materials.

## How good are Humans in Identification of LCD Recaptured Images?

We created a database of 800 photos. 400 of them were original and natural photos and the other 400 were LCD recaptured photos. To better understand human capability of detecting LCD re-captured images, we decided to make a survey-based experiment. We created a web application and invited a number of people to take part in our experiment. 47 of them have accepted our invitation and took part in experiments. The experimentation had the following process. After user signed in, he was shown a sequence of photos. There was always five seconds to decide if the photo is natural or has been recaptured from an LCD display. Natural and recaptured photos have been shown in random orders. Participants were allowed to interrupt the test process at any time, re-login, and continue anytime later. As a result, some have

been tested by, for example, just 40 images and some others for example by all 800 images.

It is also important to note that all participants were first trained by showing them a number of LCD recaptured images and a number of natural images. Recaptured pictures were taken by using three cameras and three different LCD monitors. The youngest participant was 22 years old. The oldest was 64 years old. The majority of testers were male (only 11 participants were females). Most of the participants have been either expert in the area of digital image processing or university students of various disciplines. Figure 1 shows typical examples of non-recaptured and recaptured images that have been shown to testers. Specifically, first row shows non-recaptured images. Pictures shown in the second row are recaptured from LCD.

Results obtained show that human performance in differentiating the recaptured photos from natural photos is not very good and LCD recaptured images can be effectively used to overcome authentication and image forensic systems. Specifically, the type-I error which show the rate of natural images classified as recaptured was 24.5 percent. The type-II error rate saying the rate of re-captured LCD images classified as natural images was 34 percent. Despite the fact that we spent an equal time and resources to prepare and train testers, variety in their performance was high. There were individuals with exceptional performances. On the other hand, majority of the testers exhibited poor performance. Table 1 shows performance data of 20 different individuals that participant in our survey. Shown are their performances in terms of achieved Type I error and Type II error.

## Detecting Periodic Patterns of Re-Captured LCD Screen

In this section, we will develop a method based on the theory of cyclostationarity. The method will use the periodic patterns and artifacts that are present in the recaptured images of LCD screens. Recapturing an LCD screen by using a CCD or CMOS sensor is a discrete sampling of a regular grid structure of the LCD display. The original image is first shown on the LCD screen and its regular grid and subsequently a camera is used to make a picture of it.

A visual inspection of recaptured pictures and frames of LCD display makes it possible to observe that there are often present specific periodic patterns and artifacts such as moire patterns [34,35]. They are caused by capturing the periodic sampling grid of the LCD monitor or associated aliasing. The exact appearance and intensity of the artifacts is related to several factors such as the sampling rate of the recapturing process or Color Filter Array (CFA) used in the camera [36].

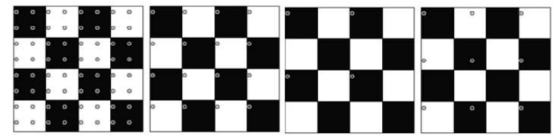
According to Nyquist frequency, the sampling rate of the recapturing process must be greater than twice the highest frequency



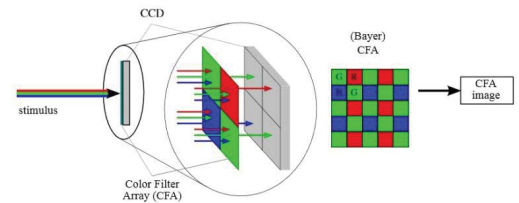
**Figure 1:** Example of typical pictures that have been used in our experiment. First row shows non-recaptured images. Pictures shown in the second row are recaptured from LCD screens.

encountered. Sampling at a lower rate results in aliasing. Figure 2 shows various types of sampling of the LCD grid. Here, the last two images show a sampling process that brings a loss of information because of an insufficient sampling rate.

As pointed out [24], CFA also plays important role in appearance and intensity of periodic patterns occurred in recaptured images. Many digital cameras are equipped with a single charge coupled device (CCD) or complementary metal oxide semiconductor (CMOS) sensor [36]. At each pixel location, only a single color sample is captured. The color images are typically obtained in conjunction with CFA (Figure 3). The most commonly used CFA is called Bayer CFA after the name of its inventor B.E. Bayer from Eastman Kodak. It consists of alternating red and green pixels on odd lines and green and blue pixels on even lines. Missing colors are computed by an interpolating process, called CFA interpolation. There are many CFA interpolation algorithms that bear into the image's different levels of spatial correlations and lead to different appearances of aliasing artifacts (bilinear, bicubic, medianbased, gradientbased, SHT, adaptive, directional filtering, etc.). For some instances of typical aliasing artifacts corresponding to different types of CFA, please see Figure 4. Here, the



**Figure 2:** Shown are various types of sampling the LCD grid. Please note that in the last two images, a loss of information occur because of an insufficient sampling rate.



**Figure 3:** Image acquisition using Bayer Color Filter Array (CFA).



**Figure 4:** Various types of occurrence of aliasing caused by three different types of CFA algorithms.

Tester ID	Type I error	Type II error
1	7.42	14.81
2	41.21	47.21
3	14	16.77
4	11.86	54.1
5	16.67	58.62
6	8.26	10.24
7	23.64	47.27
8	23.26	45.28
9	26.09	21.05
10	30.43	59.38
11	41.38	47.83
12	4.63	62.94
13	20.83	47.37
14	33.33	56.52
15	25	66.67
16	54.55	39.67
17	16.67	47.06
18	45.45	30
19	43.75	11.11
20	38.46	55.56

**Table 1:** Shown are performances of 20 different testers that participated in our experiment.

lighthouse image is shown. The image is often used for comparing various demosaicing results [37].

## Periodic Patterns

We will use the following simple, linear and stochastic model describing the recapturing process:  $f(x) = (u \cdot h)(x) + n(x)$  (Equation 1)

where  $f$ ,  $u$ ,  $h$ ,  $*$ , and  $n$  are the measured LCD screen, original LCD screen, system PSF, convolution operator, and a random variable representing the influence of noise sources that are statistically independent from the signal part of the image. In the last half a century, a lot of work has been done in the field of cyclostationarity [38]. Much of the initial work introducing and examining the use of cyclostationary models in the signal analysis was carried out by Gardner et al. [39-41]. A zero-mean signal  $f(x)$  is defined to be second order cyclostationary if its second order statistics are periodic. The autocorrelation function of  $f(x)$  can be defined as:

$$R_f(x, \delta) = E \{ f(x) f^*(x + \delta) \} \quad (\text{Equation 2})$$

Because of its periodicity in  $x$ , we can represent it in the form of a Fourier series expansion:

$$R_f(x, \delta) = \sum_{\alpha} R_f^{\alpha}(\delta) e^{j2\pi\alpha x} \quad (\text{Equation 3})$$

where  $\alpha$  is the cyclic frequency. The parameter  $R_f^{\alpha}$  is called Cyclic

Autocorrelation Function (CAF) and it is a fundamental parameter of cyclostationarity. CAF is defined as:

$$R_f^\alpha(\delta) = \lim_{x \rightarrow \infty} \frac{1}{x} \int_{-x/2}^{x/2} R_f(x, \delta) e^{-j2\pi\alpha x} dx \quad (\text{Equation 4})$$

An appropriate way of analyzing cyclostationary properties is by applying the Fourier Transform (FT) to  $R_f^\alpha$ . The result is called Spectral Correlation

### Function (SCF)

Function (SCF) and is defined as:

$$S_f^\alpha(u) = \int_{-\infty}^{\infty} R_f^\alpha(\delta) e^{-j2\pi u \delta} d\delta \quad (\text{Equation 5})$$

As we will deal with discrete signals, the discrete version of CAF should also be defined here:

$$R_f^\alpha(l) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{m=0}^{N-1} f[m] f^*[m+l] e^{-j2\pi\alpha m \Delta m} \quad (\text{Equation 6})$$

where  $N$  and  $\Delta m$  denote the number of samples of the signal and sampling interval, respectively. Equivalently, the discrete SCF can be obtained by:

$$S_f^\alpha(u) = \sum_{l=-\infty}^{\infty} R_f^\alpha(l) e^{-j2\pi u l \Delta l} \quad (\text{Equation 7})$$

CAF and SCF are analogous to the autocorrelation function and the

power spectral density function for stationary signals. When  $\alpha=0$ , the SCF can also be interpreted as the power spectral density of the signal. For other values of  $\alpha$ , SCF is the cross-spectral density between the signal and the signal shifted in frequency by  $\alpha$ . So, if the signal being analyzed exhibits cyclostationarity, the SCF will be non-zero for some  $\alpha \neq 0$ . Otherwise, only for  $\alpha=0$ , we will have non-zero values.

A cyclostationary signal has a frequency spectrum that is correlated with a shifted version of itself [40]. Based on this, in our method, we focus on detecting the traces of cyclostationarity by estimating the spectral correlation function. To estimate the SCF, we can simply use equation (7). But, due to its computational complexity, we use a more computationally effective SCF estimation method based on Fast Fourier Transform (FFT). FFT algorithm has computational complexity  $O(n \log_2 n)$ . Let's say  $f(x, y)$  is the image being analyzed and  $F(n, u)$  is a matrix containing FFT of image's rows (i.e.,  $F(1, u)$  contains the one-dimensional FFT of the first row of  $f(x, y)$ ). The SCF can be estimated in the following way:

$$S_f^\alpha(u) = \frac{1}{N} \sum_{n=0}^{N-1} F_{n,u} \bullet F_{n,u+\alpha}^* \quad (\text{Equation 8})$$

where  $*$  denotes a complex conjugate and  $N$  is the number of image's rows.

Data obtained can be combined together to create the resulting correlation map:

$$D_x = [-1 \ 1] * f(x, y) \quad D_y = \begin{bmatrix} -1 \\ 1 \end{bmatrix} * f(x, y) \quad (\text{Equation 9})$$

### Derivative filter

Periodic patterns of LCD recaptured images are often weak and not strong enough to be easily detectable using the basic cyclostationarity methods. We overcome this passing the analyzed image through a band-

pass filter. Let  $f(x, y)$  denote the image being analyzed. Then,  $D_x$  and  $D_y$  are band-passed images containing the horizontal and vertical derivative approximations:

$$D_x = [-1 \ 1] * f(x, y) \quad D_y = \begin{bmatrix} -1 \\ 1 \end{bmatrix} * f(x, y) \quad (\text{Equation 10})$$

Applying the equations (8) and (9) to  $D^n$ , where  $D^n$  is given by:

$$D^n = D_x^n + D_y^n \quad (\text{Equation 11})$$

and  $n$  denotes the order of derivative filter, results in more accurate and robust outcomes.

Application of the derivative filter to (8) results in:

$$\rho D^n(\alpha) = \sum_u |S_{D^n}^\alpha(u)| \quad (\text{Equation 12})$$

For an example of results obtained by Eq. 12 (Figure 5). Here, three digital images and obtained results are shown. Red boxes highlight the analyzed area. The image on the left (Figure 5a) is a natural and non-recaptured image. Pictures shown in middle and on right (Figure 5b and 5c) are obtained by re-capturing of an LCD screen. Distinctive peaks for pictures in middle and on right are characteristics and signifying the recapturing process. We remind that when  $\alpha=0$ , the SCF can be interpreted as the power spectral density of the signal. For other values of  $\alpha$ , SCF is the cross-spectral density between the signal and the signal shifted in frequency by  $\alpha$ . If the picture being analyzed exhibits cyclostationarity, the SCF will exhibit non-zero values for some  $\alpha \neq 0$ .

If traces of recapturing from an LCD screen is not found, only  $\alpha=0$  have

Non-zero values and no string and distinctive peak are generated. Results shown correspond to a derivative filter of order 2,  $D^2$ .

### Local blocks

Periodic patterns of recapturing have typically different intensities in various parts of the image. Thus, to successfully find traces of recapturing, we divide the image into non-overlapping blocks of  $R \times R$  pixels. We denote these blocks by  $b(x, y)$ . The developed method is always separately applied on each individual block.

### Directional analysis

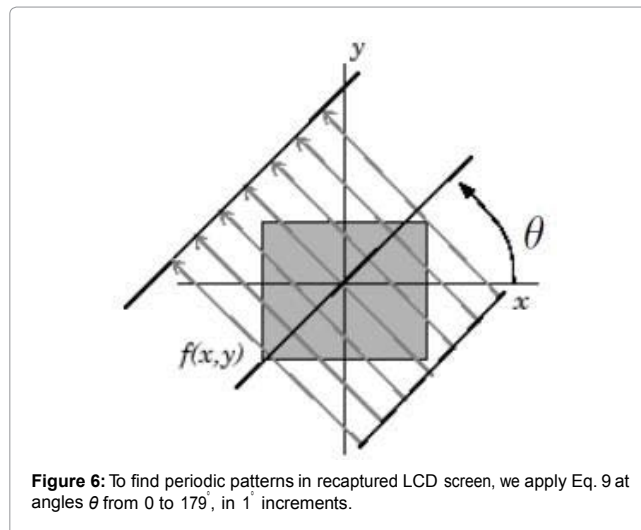
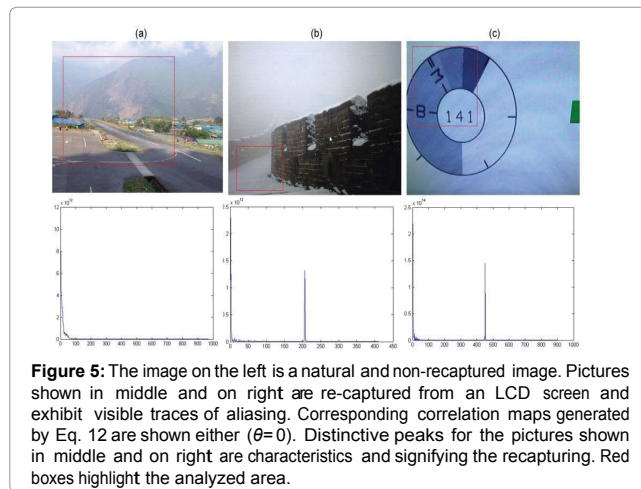
Periodic patterns corresponding to recapturing an LCD screen often are rotated and have complex spatial distributions that are caused by the rotation angle of the LCD monitor in respect to the camera that has been used to make the recapturing as well as sampling rate. To this end, we estimate the SCF and the correlation map in various directions,  $\theta$ . Specifically, we apply Eq. 8 and 9 systematically at angles  $\theta$  from  $0$  to  $179^\circ$ , in  $1^\circ$  increments (Figure 6). The new coordinate system corresponding to each  $\theta$  is obtained in the following way:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

This results in 180 vectors  $\rho\theta$ . If the investigated region has been recaptured from an LCD screen, typically, some of  $\rho\theta$  contain a specific strong peak corresponding to the cyclostationarity.

### Experiments

To validate theoretical assumptions and to measure the performance



of the developed method, we used a derivative filter of order 2,  $\rho_{D^2}(\alpha)$ :

$$\rho_{D^2}(\alpha) = \sum_u |S_{D^2}^\alpha(u)|^2 \quad [\text{Equation 13}]$$

Eq. 13 was applied to green channels of 800 tested images. Size of non-overlapping blocks  $b(x, y)$  was set to  $420 \times 420$  pixels. To automatically detect peaks corresponding to periodic artifacts of LCD recapturing, we used an automatic peak detector (P D) searching for local maximum at positions  $\alpha \neq 0$ .

The P D detects the strongest peak that is significantly larger than the local average. If the size of the found peak lies within a particular threshold, the image is classified that contains patterns of LCD recapturing, otherwise we say no traces of re-capturing have been found.

Specifically, we used a hypothesis testing approach. By assuming that the observations in maps generated by Eq. 13 are normally distributed, we formulated the null and the alternate hypothesis as follows:

$H_0$ : The peak does not correspond to recapturing

$H_a$ : The peak corresponds to a recaptured image.

Hypothesis testing was carried out in conjunction with t-statistic and a

one percent significance level. Not rejecting  $H_0$ , signifies that the peak is not a recaptured image. Type I error indicates that in one percent of the cases, we claim that an image is recaptured when it is not. In other words, we are creating a 99 percent confidence interval for the peak size to be a recaptured image.

The hypothesis testing was applied on 400 recaptured and 400 non-recaptured pictures. Results obtained have shown type I error of 3.92 percent and type II error of 8.44 percent. By empirical analysis of results, it was straightforward to observe that the quality and content of tested images play an important role in this task. For instance, detecting recaptured pictures of a very dark (almost black) content was very challenging. Moreover, as pointed out in [32], it should be noted that there are a large number of possible settings of cameras, LCD screens, surrounding lighting, etc. that have an impact on properties of recaptured data. Under specific settings, it is possible to minimize the impact of aliasing and periodic structure of the monitor grid and obtain re-captured pictures that differ very slightly from original and natural digital images. This can be achieved, for example, by optimal settings of resolution, brightness, shutter speed, capturing mode, distance between the camera and LCD, etc.

## Conclusions

In this paper, we developed a method detecting videos and pictures that have been recaptured from LCD screens. The method uses periodic patterns present in such signals. Such periodic patterns are usually caused by re-sampling of the regular pixel grid of the LCD monitor and aliasing. We used the periodicity present in statistics of such signals and automatically detected correlation in their frequency spectrum. Results obtained are promising and show that employing cyclostationarity methods can be effective in many applications in computer vision and pattern recognition including the detection of recaptured data from LCD screens.

## Acknowledgements

This work has been supported by the Czech Science Foundation under Project no. GACR 13-28462S.

## References

1. Wu M, Liu B (2002) Multimedia Data Hiding. Secaucus, NJ, USA: Springer-Verlag New York, Inc.
2. Sencar HT, Ramkumar M, Akansu AN (2004) Data Hiding Fundamentals and Applications: Content Security in Digital Multimedia. Orlando, FL, USA: Academic Press, Inc.
3. Mahdian B, Saic S (2010) "A bibliography on blind methods for identifying image forgery," Image Commun 25 389-399.
4. Dong J, Wang W, Tan T, Shi Y (2008) "Run-length and edge statistics based approach for image splicing detection," in DW08 76-87.
5. Zhang Z, Kang J, Ren Y (2008) "An effective algorithm of image splicing detection," in CSSE '08: Proceedings of the 2008 International Conference on Computer Science and Software Engineering. Washington, DC, USA: IEEE Computer Society 1035-1039.
6. Ng TT, Chang SF (2004) "A model for image splicing," in IEEE International Conference on Image Processing (ICIP), Singapore.
7. Popescu A, Farid H (2005) "Exposing digital forgeries in color filter array interpolated images," IEEE Transactions on Signal Processing 53 3948-3959.
8. Cao H, Kot AC (2009), "Accurate detection of demosaicing regularity for digital image forensics," IEEE Transactions on Information Forensics and Security 4: 899-910.
9. Huang Y, Long Y (2008) "Demosaicing recognition with applications in digital photo authentication based on a quadratic pixel correlation model," in CVPR08 1-8.

10. Swaminathan A, Wu M, Liu KJR (2008) "Hiding traces of resampling in digital images," IEEE Transactions on Information Forensics and Security 3 101-117.
11. Mahdian B, Saic S (2008) "Blind authentication using periodic properties of interpolation," IEEE Transactions on Information Forensics and Security 3 529-538.
12. Prasad S, Ramakrishnan KR (2006), "On resampling detection and its application to image tampering," in Proceedings of the IEEE International Conference on Multimedia and Exposition, Toronto, Canada 1325-1328.
13. Huang H, Guo W, Zhang Y (2008) "Detection of copy-move forgery in digital images using sift algorithm," in PACIA '08: Proceedings of the 2008 IEEE Pacific-Asia Workshop on computational Intelligence and Industrial Application. Washington, DC, USA: IEEE Computer Society 272-276.
14. Bayram S, Taha Sencar H, Memon N (2009) "An efficient and robust method for detecting copy-move forgery," in ICASSP '09, IEEE International Conference on Acoustics, Speech and Signal Processing. Washington, DC, USA: IEEE Computer Society 1053-1056.
15. Mahdian B, Saic S (2007), "Detection of copy-move forgery using a method based on blur moment invariants," Forensic science international 171: 180-189.
16. Sankar G, Zhao V, Yang YH (2009) "Feature based classification of computer graphics and real images," in ICASSP '09: Proceedings of the 2009, IEEE International Conference on Acoustics, Speech and Signal Processing. Washington, DC, USA: IEEE Computer Society 1513-1516.
17. Khanna N, Chiu GTC, Allebach JP, Delp EJ (2008) "Forensic techniques for classifying scanner, computer generated and digital camera images," in IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, USA 1653-1656.
18. Shi YQ, Chen W, Xuan G (2007) "Identifying computer graphics using hsv color model and statistical moments of characteristic functions," in ICME 1123-1126.
19. Mahdian B, Saic S (2011) "Image tampering detection using methods based on jpeg compression artifacts: A real-life experiment," in Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies, ser. ISABEL '11. New York, NY, USA: ACM, 176: 1-176.
20. Lin Z, He J, Tang X, Tang CK (2009) "Fast, automatic and fine-grained tampered jpeg image detection via dct coefficient analysis," Pattern Recogn. 42: 2492-2501.
21. Li W, Yuan Y, Yu N (2009) "Passive detection of doctored jpeg image via block artifact grid extraction," Signal Process 89: 1821-1829.
22. Kee E, Farid H (2010), "Digital image authentication from thumbnails," in Proc. SPIE, Electronic Imaging, Media Forensics and Security XII.
23. Mahdian B, Nedbal R, Saic S (2013) "Blind verification of digital image originality: A statistical approach." IEEE Transactions on Information Forensics and Security 8: 1531-1540.
24. Muammar H, Dragotti PL (2013), "An investigation into aliasing in images recaptured from an lcd monitor using a digital camera." in ICASSP 2242-2246.
25. Gao X, Qiu B, Shen J, Ng TT, Shi YQ (2011) "A smart phone image database for single image recapture detection," in Proceedings of the 9th International Conference on Digital Watermarking, ser. IWDW'10. Berlin, Heidelberg: Springer-Verlag 90-104.
26. Thongkamwitoon T, Muammar H, Dragotti PL (2014), "Robust image recapture detection using a k-svd learning approach to train dictionaries of edge profiles," In Proceeding, IEEE International Conference on Image Processing (ICIP).
27. Kose N, Dugelay JL (2012) "Classification of captured and recaptured images to detect photograph spoofing," in ICIEV 2012, IEEE/IAPR International Conference on Informatics.
28. Gao X, Ng TT, Qiu B, Chang SF (2010) "Single-view recaptured image detection based on physics-based features." in ICME. IEEE 1469-1474.
29. Scarzanella MV, Dragotti PL (2013) "Modelling radial distortion chains for video recapture detection." in MMSP. IEEE 412-417.
30. Bestagini P, Visentini-Scarzanella M, Tagliasacchi M, Dragotti P, Tubaro S (2013), "Video recapture detection based on ghosting artifact analysis," in 2013 IEEE International Conference on Image Processing (ICIP).
31. Li J, Wang Y, Tan T, Jain AK (2004) "Live face detection based on the analysis of Fourier spectra," in In Biometric Technology for Human Identification 296-303.
32. Cao H, Kot AC (2010) "Identification of recaptured photographs on lcd screens." in ICASSP. IEEE, 1790-1793.
33. Bai J, Ng TT, Gao X, Shi YQ (2010) "Is physics-based liveness detection truly possible with a single image?" in ISCAS. IEEE 3425-3428.
34. Krumm JC, Shafer SA (1991) "Sampled-grating and crossed-grating models of moire patterns from digital imaging," Optical Engineering 30: 195-206.
35. Farrow C, Shaw MM, Kim H, Juhas P, Billinge S, "The nyquist- shannon sampling theorem and the atomic pair distribution function," Physical Review B 84.
36. Lam SH, Kok CW (2001), "Demosaic: Color filter array interpolation for digital cameras," in PCM '01: Proceedings of the Second IEEE Pacific Rim Conference on Multimedia. London, UK: Springer-Verlag 1084-1089.
37. Condat L (2009) "A new random color filter array with good spectral properties." In ICIP IEEE 1613-1616.
38. Gardner WA, Napolitano A, Paura L (2006) "Cyclostationarity: Half a century of research," Signal Processing 86: 639-697.
39. Gardner WA (1987), "Spectral correlation of modulated signals: Part i-analog modulation," IEEE Transactions on Communications 35: 584-594.
40. "The spectral correlation theory of cyclostationary time-series" (1986) Signal Process 11: 13-36.
41. "Exploitation of spectral redundancy in cyclostationary signals" (1991) IEEE Signal Processing Magazine 8: 14-36.

## Submit your next manuscript and get advantages of OMICS Group submissions

### Unique features:

- User friendly/feasible website-translation of your paper to 50 world's leading languages
- Audio Version of published paper
- Digital articles to share and explore

### Special features:

- 400 Open Access Journals
- 30,000 editorial team
- 21 days rapid review process
- Quality and quick editorial, review and publication processing
- Indexing at PubMed (partial), Scopus, DOAJ, EBSCO, Index Copernicus and Google Scholar etc
- Sharing Option: Social Networking Enabled
- Authors, Reviewers and Editors rewarded with online Scientific Credits
- Better discount for your subsequent articles

Submit your manuscript at: <http://www.omicsonline.org/submission>

**Citation:** Mahdian B, Amsky AN, Saic S (2015) Detecting Cyclostationarity in Re-Captured LCD Screens. J Forensic Res 6: 294. doi: [10.4172/21577145.1000294](https://doi.org/10.4172/21577145.1000294)

# Determination of Stop-Criterion for Incremental Methods Constructing Camera Sensor Fingerprint

Babak Mahdian<sup>(\*)</sup>, Adam Novozámský,  
and Stanislav Saic

Institute of Information Theory and Automation of ASCR,  
Prague, Czech Republic  
{mahdian,novozamsky,ssaic}@utia.cas.cz

**Abstract.** This paper aims to find the minimum sample size of the camera reference image set that is needed to build a sensor fingerprint of a high performance. Today's methods for building sensor fingerprints do rely on having a sufficient number of camera reference images. But, there is no clear answer to the question of how many camera reference images are really needed? In this paper, we will analyze and find out how to determine the minimum needed number of reference images to remove the mentioned uncertainty. We will introduce a quantitative measure (a stop-criterion) stating how many photos should be used to create a high-performance sensor fingerprint. This stop-criterion will directly reflect the confidence level that we would like to achieve. By considering that the number of digital images used to construct the camera sensor fingerprint can have a direct impact on performance of the sensor fingerprint, it is apparent that this, so far underestimated, topic is of major importance.

**Keywords:** Image ballistics · Source camera verification · Pattern noise · PRNU · Fingerprint performance · Laplace distribution

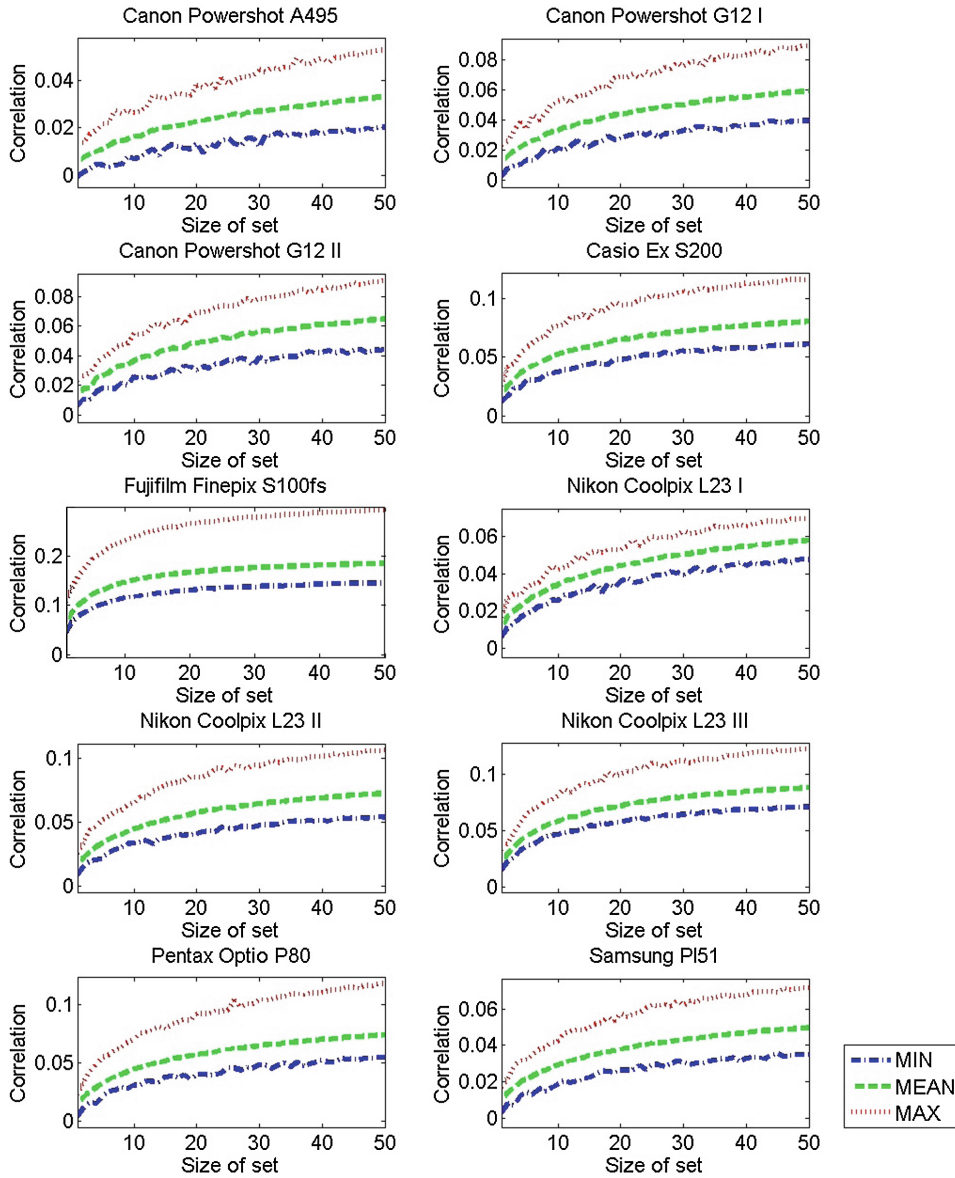
## 1 Introduction

Generally, there are two essential tasks in forensics analysis of digital images and videos: their integrity verification (genuineness analysis) and ballistics analyzes. In this paper we will deal with image (video) ballistics which does address the problem of linking digital images (videos) under investigation to the exact source imaging device that has been used to capture photos (videos) under investigation. Since image ballistics makes possible to differentiate between source cameras of the same make and model, it became especially useful in the forensic, law enforcement, insurance, and media industries.

Although past research was mainly focused on data hiding and digital watermarking approaches [1–3] to perform digital image integrity verification and image ballistics, today there is a relatively new approach called passive one which

does not need embedding any secondary data into the image [4]. In contrast to active methods, the passive approach does not need any prior information about the image being analyzed. There have been methods developed to detect image splicing [5,6], traces of non-consistencies in color filter array interpolation [7], traces of geometric transformations, [8], cloning [9], computer graphics generated photos [10], JPEG compression inconsistencies [11], etc. Typically, pointed out methods are based on the fact that digital image editing brings specific detectable statistical changes into the image.

In the image ballistics area, methods mainly focused on camera sensor noise and systematic artifacts that are brought into the image [12–18]. These artifacts have been used to link a digital image to its exact acquisition device.



**Fig. 1.** Performance of camera sensor fingerprint constructed using camera reference sets of different sizes and 100 test images for each camera.

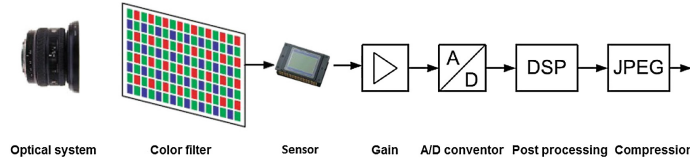
### 1.1 Motivation

When linking a digital image to an exact camera (or a video signal to camcorder), typically the following procedure is used. First, the camera sensor fingerprint is constructed [12,13]. Second, the constructed fingerprint and image under investigation are matched (usually through a correlation measurement). This indicates if the digital image has been captured by this camera. The sensor fingerprint is constructed incrementally by using many camera reference images. Camera reference images are recommended to be of a uniformly illuminated surface. Usually, an edge-preserving denoising filter is applied on camera reference images. Residuals of digital images and their denoised versions are put together (e.g., by averaging) to construct the basic version of sensor fingerprint.

The problem is how many photos should be used to form the camera reference images so there will be a high confidence that the constructed fingerprint is of a high performance? Is the optimal size of this set 10, 50, or even 250? The topic is of major importance. The reason is that number of digital images used to construct the camera sensor fingerprint, typically, has a direct impact on performance of the sensor fingerprint. Insufficient number of reference images cause a poor performance of the fingerprint. To this end, most authors rather recommended to employ a fixed and higher number of reference images to be safe in terms of having a good performing fingerprint (in published literature we have, typically, observed recommended sizes of reference images ranging from 30 to 150).

To remove the uncertainty about the size of sets of camera reference images, we will introduce a quantitative measure determining how many photos should be used to create the sensor fingerprint to have a high performing fingerprint. In other words, we are going to search for the optimal number of reference images that will reflect our the confidence level and accuracy we want to achieve. To address the problem we will search for a stop-criterion stating that no more images are needed to be added to the set of camera reference images.

Before going on, we also explicitly define what is a fingerprint of good (high) or poor performance. A fingerprint with a good performance is such a fingerprint that enables a successful recognition of the exact source cameras when inspecting photos of various scenes, lighting conditions, etc. When a non-sufficient number of images are used to create the sensor fingerprint, the measured fingerprint is of poor performance (often random noise components dominate in there) and hence the image source verification task generate weak results. By weak results we mean lower rate of true positives. Figure 1 demonstrates performance of 10 different camera sensor fingerprints constructed by  $1 \dots 50$  reference images of uniformly illuminated surface (a white paper). The figure demonstrates obtained correlation (obtained by using Eq. 3) between 100 test images (natural images captured by same cameras) and associated camera fingerprints (minimal, maximal and mean values of obtained correlation values are shown). Apparently size of camera reference sets have a direct impact on results obtained. For the sake of completeness, we point out that false positive states for mistakenly pinpointing



**Fig. 2.** A typical digital camera system.

the source camera. By true positive we mean correctly pinpointing a digital image to the source camera.

## 2 Basic Notations and Preliminaries

A typical camera consists of several different components (see Fig. 2). As pointed out in [19], the core of every digital camera is the imaging sensor. The sensor (e.g., CCD or CMOS) is consisted on small elements called pixels that collect photons and convert them into voltages that are subsequently sampled to a digital signal in an A/D converter. Generally, before the light from the scene which is being photographed reaches the sensor it also passes through the camera lenses, an antialiasing (blurring) filter, and then through a color filter array (CFA). The CFA is a mosaic of tiny color filters placed over the pixel of an image sensor to capture color information. Color filters are needed because typical consumer cameras only have one sensor which cannot separate color information. Most commonly, Bayer color filter is used.

The resulting signal is then further processed using color correction and white balance adjustment. Additional processing includes gamma correction to adjust for the linear response of the imaging sensor, noise reduction, and filtering operations to visually enhance the final image. Finally, the digital image might be compressed stored and stored in a specific image format like JPEG.

What is important in terms of forensic analyzes of digital images is that different components of camera leave different kind of artifacts or fingerprints useful for integrity verification of photos or ballistics analysis. Typically, artifacts (fingerprints) left by CFA, post processing, and compression parts are in common for cameras of same make and model. In other words, assuming that we know their value and behavior for a particular camera make and model and based on the fact that digital image editing (e.g., photoshopping) change these values (fingerprints), they can be employed for verification of the originality of digital images.

On the other hand, each camera has its own unique sensor which consists of millions of pixels each of unique properties. Hence, if we are able to find kind of information brought into image by the sensor and which will remain stable and present in all images captured by that sensor and cannot be found in no image captured by any other sensor, then we can call it fingerprint of that sensor or camera. Such a camera sensor fingerprint can be employed to link digital images to particular digital cameras which captured them.

## 2.1 Sensor as a Camera Fingerprint

Image sensors suffer from several fundamental and technology related imperfections resulting in their performance limitations and noise. As pointed out in [19], if we take a picture of an absolutely evenly lit scene, the resulting digital image will still exhibit small changes in intensity among individual pixels which is partly because of pattern noise, readout noise or shot noise. While readout noise or shot noise are random components, the pattern noise is deterministic and remain approximately the same if multiple pictures of the same scene are taken. As a result, pattern noise can be the fingerprint of sensors which we are searching for.

Pattern Noise (PN) is consisted of two components called Fixed Pattern Noise (FPN) and photo response non-uniformity (PRNU). FPN is independent of pixel signal, it is an additive noise, and some high-end consumer cameras can suppress it. The FPN also depends on exposure and temperature. PRNU is formed by varying pixel dimensions and inhomogeneities in silicon resulting in pixel output variations. It is a multiplicative noise. Moreover, it is not dependent on temperature and seems to be stable over time. The values of PRNU noise increases with the signal level (it is more visible in pixels showing light scenes). In other words, in very dark areas PRNU noise is suppressed. Moreover, PRNU is not present in completely saturated areas of an image. Thus, such images should be ignored when searching for PRNU noise.

There has not been performed a lot of studies analyzing the PRNU noise in deeper details. Despite this, it has been shown that it has a dominant presence in the pattern noise component found in digital images. This made possible Fridrich et al. [12, 13] to employ PRNU noise to identify exact source cameras. In other words, PRNU noise is employed as the fingerprint of camera sensors. Generally, it can be claimed that state-of-the-art source identification methods are mostly based on methods proposed by Fridrich et al. (e.g., [12, 13, 20, 21]). There have been published some additional papers by others authors (e.g., [14–18]) aiming to improve accuracy of results. Typically, they brought modifications to the original paper of Fridrich et al. [12, 13] based on some new theoretical or empirical findings. Nonetheless, the key concept of how to measure sensor’s fingerprint has remained unchanged.

## 2.2 Modeling and Extracting PRNU

Let us model the image acquisition process in the following way:

$$I_{i,j} = I_{i,j}^o + I_{i,j}^o \cdot \Gamma_{i,j} + \Upsilon_{i,j} \quad (1)$$

Here,  $I_{i,j}$  denotes the image pixel at position  $(i, j)$  produced by the camera,  $I_{i,j}^o$  denotes the noise-free image (perfect image of the scene),  $\Gamma_{i,j}$  denotes PRNU noise and  $\Upsilon_{i,j}$  stands for all additive or negligible noise components.

Following the approach proposed by [12, 13], the PRNU component is estimated in the following way. For a given camera, PRNU noise is estimated by

averaging multiple images  $I_k$ ,  $k = 1, \dots, N$  captured by this camera. The process is sped up by suppressing the scene content from the image prior to averaging. This is achieved by using a denoising filter  $\mathcal{F}$  and averaging the noise residuals instead. We will denote residuals by  $\hat{I}_k$  (i.e.,  $\hat{I}_k = I_k - \mathcal{F}(I_k)$ ). In other words, deterministic components of sensor noise of the camera  $C$  are computed in the following way:

$$\Gamma_{sensor} = \frac{1}{N} \sum_{k=1}^N \hat{I}_k = \frac{1}{N} \sum_{k=1}^N I_k - \mathcal{F}(I_k) \quad (2)$$

Alternatively, for example, maximum likelihood estimation (MLE) instead of simple averaging can be employed.

To reduce the false positive rate, sensor fingerprint are enhanced by Wiener filtering in the frequency domain (e.g., to reduce JPEG compression artifacts) and linear pattern removal through zero-mean operation (e.g., to remove traces of CFA interpolation) [12]. Pointed out  $\Gamma_{sensor}$  is the basic version of sensor fingerprint of camera. To achieve accurate results and minimize the false positive rate, it is necessary to perform additional frequency filtering, fingerprint enhancement and correction, suppressing dominant traces of camera embedded software, filtering JPEG artifacts, etc. Without such a correction, typically, a high rate of false positives is obtained (because of camera operations such as gamma correction, CFA interpolation, color enhancement, geometric deformation corrections, compression, additional embedded camera software functionalities, etc.) This part often depends on specific camera brands under investigation.

Linking of a digital image to an exact camera is carried out by performing a similarity measure of two sensor fingerprints. One is obtained from the image under investigation and second from the set of camera reference images. This can be carried out, for example, by employing a simple correlation measure. Having available two different sensor fingerprints  $\Gamma_{s_1}$  and  $\Gamma_{s_2}$ , we measure their similarity by employing a normalized correlation:

$$corr(\Gamma_{s_1}, \Gamma_{s_2}) = \frac{(\Gamma_{s_1} - \overline{\Gamma_{s_1}}) \odot (\Gamma_{s_2} - \overline{\Gamma_{s_2}})}{(\|\Gamma_{s_1} - \overline{\Gamma_{s_1}}\|) \cdot (\|\Gamma_{s_2} - \overline{\Gamma_{s_2}}\|)} \quad (3)$$

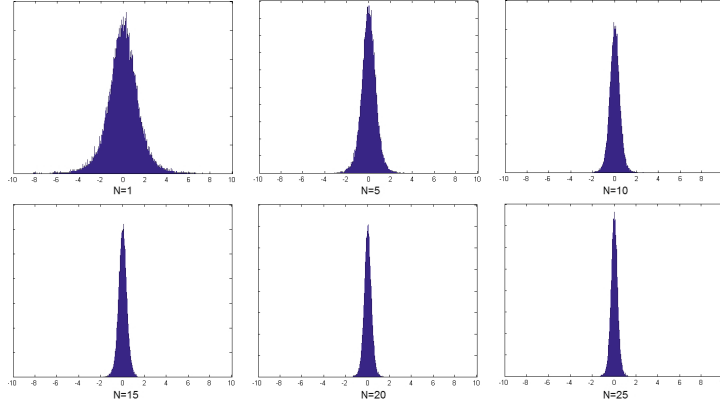
where  $\overline{X}$  denotes mean of the vector  $X$ ,  $\odot$  stands for dot product of vectors defined as  $X \odot Y = \sum_{k=1}^N X(k)Y(k)$  and  $\|X\|$  denotes  $L_2$  norm of  $X$  defined as  $\|X\| = \sqrt{X \odot X}$ .

There has been carried out studies about the specific choice and effectiveness of denoising filters (e.g., [14]). It is important to note that there is no general perfect denoising filter. All of them have their advantages and disadvantages. Moreover, it is interesting to note that when applying the proposed PRNU estimation method on a larger set of digital images or when analyzing digital video signals consisted of thousands of individual frames, the computational time becomes a drawback of the method. It has been shown that the computational time of the method can effectively be enhanced by using GPU-accelerated version of the algorithm. For example, in [22, 23] a parallel CUDA implementation of  $\Gamma_{sensor}$  has been built achieving remarkable speedup in fingerprint computation (up to 5–6 times).

### 3 Laplacian Distributed Residuals

As pointed out in last section, for a given camera  $C$ , deterministic components of sensor noise can be estimated by averaging multiple images captured by this camera,  $I_k$ ,  $k = 1, \dots, N$ . The process is sped up by suppressing the scene content from the image prior to averaging by using a denoising filter  $\mathcal{F}$  and averaging the noise residuals  $\hat{I}_k$  instead.

Apparently samples of residuals,  $\hat{I}_k$ , and their corresponding averaged versions  $\frac{1}{N} \sum_{k=1}^N \hat{I}_k$  are the key information forming the sensor fingerprint of camera,  $\Gamma_C$ . Let us first to find an appropriate form for the probability density function (p.d.f.) of the distribution of residual values so that they can be efficiently modeled. Figure 3 demonstrates the histogram of residuals  $\frac{1}{N} \sum_{k=1}^N \hat{I}_k$ , obtained using a typical set of reference images of sizes  $N = 1, 5, 10, 15, 20, 25$ .



**Fig. 3.** Distribution of averaged residuals  $\Gamma_C$  constructed using a typical camera reference image set of different sizes  $N = 1, 5, 10, 15, 20, 25$ .

These figures demonstrate that the Laplacian p.d.f. fits the observed distribution well.  $\Gamma_C$  has a Laplace  $(\mu, b)$  distribution if its probability density function is

$$f(\Gamma_{C_{i,j}} | \mu, b) = \frac{1}{2b} \exp\left(-\frac{|\Gamma_{C_{i,j}} - \mu|}{b}\right) \quad (4)$$

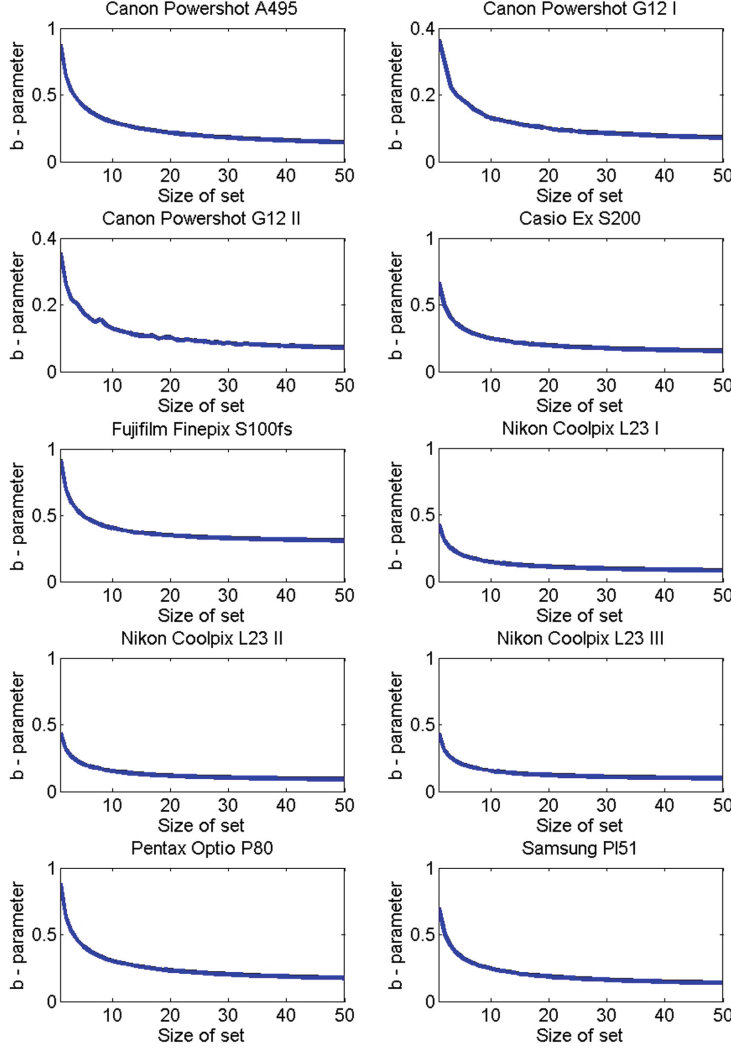
where  $\mu$  is a location parameter and  $b \geq 0$  is sometimes referred to as the diversity.

To estimate parameters of the Laplace distribution, maximum likelihood estimator is used. Maximum likelihood estimator of  $b$  can be obtained by:

$$\hat{b} = \frac{1}{M} \sum_{i=1}^M |\Gamma_{C_{i,j}} - \hat{\mu}| \quad (5)$$

Having Laplacian-distributed residuals,  $\hat{I}_k$ , we easily can estimate parameters the of associated p.d.f. Specifically, we focus on parameter  $b$  and will analyze its

behavior during computation process of the sensor fingerprint. Figure 4 demonstrates values of  $b$  for different number camera reference images of 10 different cameras. Apparently,  $b$  follows a descending trend. Specifically, we can see that  $b$  descends as the number of camera reference images grows. It is important to note that the descending trend is steep in the beginning. On the other hand,  $b$  becomes almost stable for higher number of images.



**Fig. 4.** Estimated parameter  $b$  for 10 different sensor fingerprints constructed using camera reference image sets of sizes  $N = 1 \dots 50$ .

#### 4 Determination of Stop-Criterion for Size of Reference Images

The uncertainty which is addressed in this paper is about the needed number of reference images,  $N$ , that is needed to construct a sensor fingerprint,  $\Gamma_{sensor}$ , of high performance. As pointed out previously, in literature, it is often pointed out that  $N \rightarrow \infty$  brings a more accurate sensor fingerprint and suppressed  $\mathcal{Y}$ .

In last section we introduced the parameter  $b$  that is based on Laplace distribution modeling of residuals. Moreover, we have shown that  $b$  has a specific behavior and descends as the number of camera reference images grows. It has been shown that the descending trend of  $b$  is steep in the beginning. On the other hand,  $b$  becomes almost stable for higher number of images.

The question is what is the relation between  $b$  and the performance of the sensor fingerprint and how can we employ  $b$  to predict the future performance level of the fingerprint in terms of true positives? Here, we will use a differential operator to quantify the behavior of  $b$ . In other words, having  $\Gamma_C = \frac{1}{N} \sum_{k=1}^N \hat{I}_k$ , we will measure the rate at which the value of the  $b$  changes with respect to change of  $k$ :

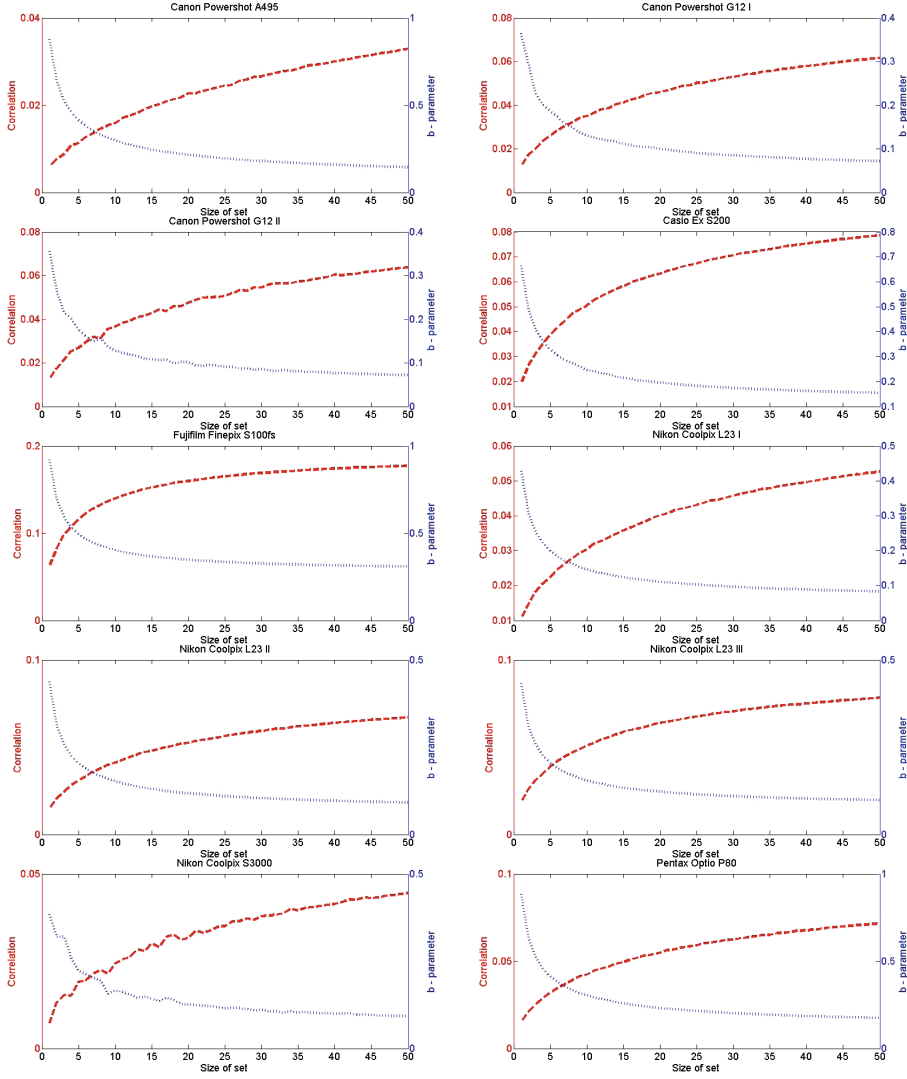
$$\Delta b_k = b_{k+1} - b_k \quad (6)$$

To create a stop-criterion, we collected 25 different cameras and for each of them created a camera reference image sets of 50 images and a test image sets of 100. We constructed 50 sensor fingerprints  $\Gamma_C$  by using  $1 \cdots 50$  reference images per camera. Reference photos have been selected randomly. Having 50 different sensor fingerprints constructed using a different size of reference image sets, we carried out an image ballistics test using Eq. 3 that calculated the true positive rate for all test images. A global threshold has been employed for the classification part of this part. At the same time we also measured Eq. 6 for all camera fingerprints (to remove local outliers, a low-pass filter always have always been applied on  $\Delta b$ ). Having available 50 different true positive rates as well as 50 values of  $b$  for each camera sensor fingerprint, we analyzed their relation and empirically gained an optimal  $b$  for different rates of true positives. Efficiency that can be obtained by using  $b$  as the stop-criterion is shown the next section. In this study, we selected the false positive rate to be 0.1 percent. For the sake of simplicity, for the parameter searching and experimental part, there only have been chosen cameras that are distinguishable using the basic version of sensor fingerprint enhanced by Wiener filtering in the frequency domain (e.g., to reduce JPEG compression artifacts) and linear pattern removal through zero-mean operation as recommended in [12].

Figure 5 demonstrates a portion of results of our analysis. Specifically, shown is performance of fingerprint constructed by sets of reference images of sizes  $N = 1 \cdots 50$ . Also shown is the associated and estimated  $b$ .

## 5 Experimental Results

Table 1 demonstrates efficiency of using  $\Delta b$  for 10 test camera sensor fingerprints. These cameras have not been used in the process of determination of optimal  $\Delta b$ . We have selected our true positive rate to be 99.99 percent with having false positive rate of 0.01 percent. Considering these desired true and false positive rates we have computed the optimal  $\Delta b$ . As mentioned in last section, this stop-criterion has been calculated using a 25 different cameras and associated sets of reference images of sizes  $N = 1 \cdots 50$ . In our case, optimal  $\Delta b$  was 0.0050. For each test camera, Table 1 shows gained true positive rate and associated size



**Fig. 5.** Shown is performance of fingerprint constructed by sets of reference images of sizes  $N = 1 \cdots 50$ . Also shown is the associated and estimated  $b$ .

**Table 1.** Shown is performance (%) of sensor fingerprints based on different values of  $\Delta b$ . Shown is also associated number of reference images used to construct the sensor fingerprint (shown in square brackets).

$\Delta b$	0.0400	0.0300	0.0200	0.0100	0.0050	0.0040	0.0030
canon-powershot-g12-III	73 [3]	84 [4]	99 [6]	100 [9]	100 [14]	100 [16]	100 [19]
canon-powershot-g12-IV	75 [2]	97 [4]	100 [7]	100 [10]	100 [11]	100 [13]	100 [22]
fujifilm-finepix-s100fs-II	100 [4]	100 [5]	100 [6]	100 [10]	100 [14]	100 [21]	100 [25]
nikon-coolpix-l23-IV	81 [3]	93 [5]	97 [6]	100 [8]	100 [11]	100 [15]	100 [19]
nikon-coolpix-l23-V	100 [4]	100 [5]	100 [5]	100 [7]	100 [11]	100 [17]	100 [24]
nikon-coolpix-l23-VI	100 [3]	100 [4]	100 [6]	100 [7]	100 [12]	100 [14]	100 [21]
pentax-optio-p80-II	97 [4]	100 [5]	100 [7]	100 [11]	100 [18]	100 [22]	100 [27]
iphone-3GS	18 [5]	28 [6]	28 [7]	62 [12]	94 [20]	100 [27]	100 [35]
iphone-4s-I	86 [4]	98 [5]	100 [6]	100 [10]	100 [15]	100 [20]	100 [23]
iphone-4s-II	95 [4]	100 [6]	100 [8]	100 [9]	100 [12]	100 [21]	100 [28]

of camera reference image set (shown in square brackets). In this experiment, we gained 0 percent false positive rate.

Table 1 demonstrates gained true positive rates for 100 test images per camera. Moreover, size of camera reference images used to construct the camera sensor fingerprint is shown either. For the sake of completeness, we also show different values of  $\Delta b$  and associated results obtained.

## 6 Discussion and Conclusion

In this paper we addressed the problem of uncertainty of how many reference images should be used to construct a high performance camera sensor fingerprint. Typically, papers dealing with construction of sensor fingerprints proposed to incrementally use about 40–50 images of a uniformly illuminated surfaces. Some others simply recommended to use as much as possible.

In last sections, we have introduced a quantitative measure stating how many photos should be used. We searched for an the minimal number of camera reference images that will directly reflect the confidence and accuracy level we want to achieve. To address the problem, we introduced a stop-criterion that can determine if more images are needed to be added to the set of reference images to get the desired true positive rate.

It also has been shown that a small number of reference images available for construction a sensor fingerprint is not always a limiting factor for constructing a well performing fingerprint. It should be noted that employing camera reference image sets of  $N$  when  $N \rightarrow \infty$  does not necessarily convert to a perfect sensor fingerprint. By a perfect sensor fingerprint we mean a signal that only and only consists of deterministic noise components unique for each sensor. Digital images captured by today's consumer cameras and smartphones suffer from a set of systematic and non-systematic imperfections and enhancements (sensor noise, gamma correction, CFA interpolation, color enhancement, geometric deformation corrections, and a number of additional embedded camera software functionalities) that bring a number of correlated and uncorrelated artifacts into digital images which cannot be overcome using  $N$ ,  $N \rightarrow \infty$ , number of camera reference image.

**Acknowledgments.** This work has been supported by the Czech Science Foundation under Project no. GACR 13-28462S.

## References

1. Sencar, H.T., Ramkumar, M., Akansu, A.N.: Data Hiding Fundamentals and Applications: Content Security in Digital Multimedia. Academic Press Inc., Orlando (2004)
2. Arnold, M., Schmucker, M., Wolthusen, S.D.: Techniques and Applications of Digital Watermarking and Content Protection. Artech House Inc., Norwood (2003)
3. Nikolaidis, N., Pitas, I.: Robust image watermarking in the spatial domain. Signal Process. **66**(3), 385–403 (1998)

4. Mahdian, B., Saic, S.: A bibliography on blind methods for identifying image forgery. *Image Commun.* **25**(6), 389–399 (2010)
5. Ng, T.-T., Tsui, M.-P.: Camera response function signature for digital forensics - part i: theory and data selection. In: *IEEE Workshop on Information Forensics and Security*, December 2009, pp. 156–160 (2009)
6. Lint, Z., Wang, R., Tang, X., Shum, H.-Y.: Detecting doctored images using camera response normality and consistency. In: *CVPR 2005: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005) - Volume 1*, Washington, pp. 1087–1092. IEEE Computer Society (2005)
7. Popescu, A., Farid, H.: Exposing digital forgeries in color filter array interpolated images. *IEEE Trans. Signal Process.* **53**(10), 3948–3959 (2005). [www.cs.dartmouth.edu/farid/publications/sp05a.html](http://www.cs.dartmouth.edu/farid/publications/sp05a.html). [Online]
8. Mahdian, B., Saic, S.: Blind authentication using periodic properties of interpolation. *IEEE Trans. Inf. Forensics Secur.* **3**(3), 529–538 (2008)
9. Mahdian, B., Saic, S.: Detection of copy-move forgery using a method based on blur moment invariants. *Forensic Sci. Int.* **17**(2–3), 180–189 (2007)
10. Dirik, A.E., Bayram, S., Sencar, H.T., Memon, N.: New features to identify computer generated images. In: *IEEE International Conference on Image Processing, ICIP 2007*, vol. 4, pp. 433–436 (2007)
11. Fridrich, J., Pevny, T.: Detection of double-compression for applications in steganography. *IEEE Trans. Inf. Sec. Forensics* **3**(2), 247–258 (2008)
12. Chen, M., Goljan, M., Lukas, J.: Determining image origin and integrity using sensor noise. *IEEE Trans. Inf. Forensics Secur.* **3**(1), 74–90 (2008)
13. Lukas, J., Fridrich, J., Goljan, M.: Digital camera identification from sensor pattern noise. *IEEE Trans. Inf. Forensics Secur.* **1**(2), 205–214 (2006)
14. Amerini, I., Caldelli, R., Cappellini, V., Picchioni, F., Piva, A.: Analysis of denoising filters for photo response non uniformity noise extraction in source camera identification. In: *Proceedings of the 16th International Conference on Digital Signal Processing, Series DSP 2009*, Piscataway, pp. 511–517. IEEE Press (2009). Available: <http://dl.acm.org/citation.cfm?id=1700307.1700392>. [Online]
15. Alles, E.J., Geradts, Z.J.M.H., Veenman, C.J.: Source camera identification for heavily JPEG compressed low resolution still images. *J. Forensic Sci.* **54**(3), pp. 628–638 (2009). Available: <http://www.science.uva.nl/research/publications/2009/AllesJFS2009>. [Online]
16. Yongjian Hu, C.J., Yu, B.: Source camera identification using large components of sensor pattern noise. In: *2nd International Conference on Computer Science and Its Applications, CSA 2009*, Jeju Island, Korea (2009)
17. Li, Y., Li, C.-T.: Decomposed photo response non-uniformity for digital forensic analysis. In: *Sorell, M. (ed.) e-Forensics 2009*. LNCS, vol. 8, pp. 166–172. Springer, Heidelberg (2009)
18. Hu, Y., Jian, C., Li, C.-T.: Using improved imaging sensor pattern noise for source camera identification. In: *ICME*, pp. 1481–1486 (2010)
19. Lukas, J., Fridrich, J., Goljan, M.: Detecting digital image forgeries using sensor pattern noise. In: *Proceedings of the SPIE, West*, p. 2006 (2006)
20. Chen, M., Fridrich, J., Goljan, M., Luk, J.: Source digital camcorder identification using sensor photo-response nonuniformity. In: *Proceedings of SPIE Electronic Imaging, Photonics West* (2007)
21. Chen, M., Fridrich, J., Goljan, M.: Digital imaging sensor identification (further study). In: *Delp, E.J., III; Wong, P. W. (ed.) Security, Steganography, and Watermarking of Multimedia Contents IX*. *Proceedings of the SPIE*, vol. 6505 (2007)

22. Williams, D., Codreanu, V., Yang, P., Liu, B., Dong, F., Yasar, B., Mahdian, B., Chiarini, A., Zhao, X., Roerdink, J.: Evaluation of autoparallelization toolkits for commodity graphics hardware. In: 10th International Conference on Parallel Processing and Applied Mathematics. Springer, Warsaw, Poland (2013)(to appear)
23. Williams, D., Codreanu, V., Roerdink, J.B., Yang, P., Liu, B., Dong, F., Chiarini, A.: Accelerating colonic polyp detection using commodity graphics hardware. In: Proceedings of the International Conference on Computer Medical Applications, Sousse, Tunisia, pp. 1–6 (2013)



# Copy-move forgery detection using JPEG compression model

A. Novozámský<sup>a,\*</sup>, M. Šorel<sup>a</sup>

<sup>a</sup>*Department of Image Processing, Institute of Information Theory and Automation, Czech  
Academy of Sciences, Pod Vodárenskou věží 4, CZ-182 08, Prague 8*

---

## Abstract

The copy-move forgery, based on copying an object and pasting in another location of the same image, is a common way to manipulate image content. In this paper, we address the problem of copy-move forgery detection in JPEG images. The main problem with JPEG compression is that the same pixels, after moving to a different position and storing in the JPEG format, have different values. The majority of existing algorithms is based on matching pairs of similar patches, which generates many false matches. In many cases they cannot be eliminated by postprocessing, causing the failure of detection. To overcome this problem, we derive a JPEG-based constraint that any pair of patches must satisfy to be considered a valid candidate and propose an efficient algorithm to verify the constraint. The constraint can be integrated into most existing methods. Experiments show significant improvement of detection, especially for difficult cases, such as small objects, object covered by textureless areas and repeated patterns.

*Keywords:* image forensics, copy-move forgery detection, image tampering, JPEG, quantization constraint set

---

## 1. Introduction

The integrity of visual data is important for the credibility of news media and especially when used as an evidence in court or during criminal investigation. In the times of analog recording and classical photography any tampering was considered difficult and time consuming but the availability of easy-to-use image processing software made the manipulation of digital content extremely simple. Image forensics, a branch of forensic data analysis, evolved as a scientific means to verify the source of image data and detect potential modifications.

One of the most common image modifications is the *copy-move* forgery, based on copying an object and pasting in another location of the same image.

---

\*Corresponding author

*Email addresses:* `novozamsky@utia.cas.cz` (A. Novozámský), `sorel@utia.cas.cz` (M. Šorel)

Transition between the inserted object and original contents is often masked by various retouching tools. Copying from the same image keeps statistical properties of the image such as the noise, contrast and color, which makes detection more difficult. On the other hand, reusing the same object in one image can be detected and is what is looked for by the majority of copy-move forgery detection (CMFD) techniques.

While the copy-move forgery is relatively easy to detect in images stored in lossless formats, such as PNG, GIF, and TIFF, detection in JPEG images is complicated by the fact that the same pixels, after moving to a different position and storing in the JPEG format, have different values. For this reason, when looking for copy-move candidates, existing methods usually consider all patches that are in a sense similar. Irrespective of how the similarity is measured, the problem arises that the criterion used is always a compromise between detecting all true candidates and getting a reasonable number of false positives, i. e. patches that actually were not copied but must be considered valid candidates.

In this article, we analyze the problem whether it is possible to reduce the number of false positives using the exact mechanism of JPEG compression. In other words, whether it is possible to say something about how originally identical patches can differ under compression.

The solution we present in this article is based on the observation that all original images that could have resulted in the coefficients stored in the JPEG file we analyze, form a convex set, in the field of digital image restoration known as the quantization constraint set [34]. This set can be specified separately for each JPEG block, typically of size  $8 \times 8$  pixels. For grayscale images, the set has a form of a multi-dimensional box (in general orientation) centered in the values we get by ordinary JPEG decompression. The dimensions of this box are given by the values from the quantization table stored in every JPEG file during compression. Therefore, since the patch we are analyzing is present at two different places, the original question can be reformulated as verifying whether the intersection of two sets is empty or not. Is there an original patch that would be compressed to two different sets of coefficients stored in two different positions in the JPEG file? In this paper, we call this condition the *JPEG copy-move constraint* and show an efficient method how its validity can be verified. We show that the idea of using the JPEG copy-move constraint is compatible with most existing methods. It helps especially in images with repeated motives and objects retouched by an indistinct texture, which are common situations extremely difficult for all methods we can find in literature.

## 2. Related work

Copy-move forgery detection was first proposed by Fridrich et al. [15], which inspired a large amount of research in the same direction. The number of papers has been increasing every year and CMFD has become one of the most active topics in image forensics. In this section, we give a short overview of main approaches, for a more complete treatment, we refer readers to [9, 28] comparing

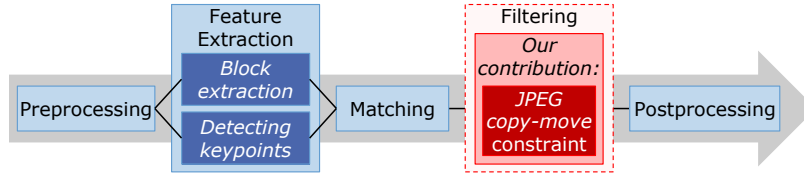


Figure 1: Typical processing pipeline of copy-move forgery detection algorithms. The main contribution of this paper is filtering using the JPEG copy-move constraint, which increases sensitivity and reduces significantly the number of false matches.

many popular algorithms. Readers interested in image forensics in general can start with [24] and tutorials available on the web pages of Hany Farid.

The typical processing pipeline used by most CMFD methods is shown in Fig. 1 (derived from Fig. 2 in [9]). The preprocessing phase includes decompression of image files if stored in a compressed format and for methods working on grayscale images also conversion to grayscale.

Based on the second step (the second block from left in Fig. 1), CMFD methods can be divided into two large groups, block-based and keypoint-based, depending on the mechanism used to find the candidates for matching. The vast majority of them is in the first group, including [15]. In this approach, algorithms look for similar block obtained by partitioning the image into overlapping blocks. The reason, why it is usually impractical to look for exactly the same values (exact match in [15]) is that values can be damaged by JPEG compression or additional retouching operations. Methods either work directly with pixels or compute features transforming blocks into a representation of lower-dimension or having convenient invariant properties. For example, [15] used weighted DCT transform coefficients and Bashar et al. [3] the Daubechies four tap wavelet transform.

An important reason for using features is to achieve invariance with respect to some transforms. While the basic copy-move detection assumes that object is unchanged and in the same orientation, in practice it is possible that before being pasted, it was rotated, scaled or even blurred. To ensure the invariance to rotation, Wang et al. [32] used a circle block model, Bayram et al. [6] added a scale invariance by features obtained from the Fourier-Mellin transform. Several authors worked with moment invariants. Mahdian and Saic [21] suggested 24 blur invariants, Ryu et al. [27] used the Zernike moments.

In the matching phase (third phase in Fig. 1), algorithms take the blocks or their features stored in a feature matrix and look for similar entries. Since considering all pairs of blocks would be extremely time-consuming, methods save time in various ways. One possibility is decreasing the number of features, which reduces the dimension of space we are searching. Popescu et al. [25] reduced the feature matching space by the PCA applied on blocks. Bashar et al. [3] modified their algorithm to use the kernel PCA. Huang et al. [17] simply shortened the feature vector to its first quarter. A complementary way to reduce computational complexity is using an efficient data structure for the approximate nearest-neighbor search. In the original paper [15], similar blocks

were found by lexicographically sorting the rows in the feature matrix and comparing the adjacent rows, which was used by many followers. A more reliable efficient data structure are kd-trees, mostly used in the best-bin-first variant [7], which is applicable in both block and keypoint-based algorithms. Comparison [9] uses multiple randomized kd-trees [22]. Another fast iterative randomized technique for computing the approximate nearest-neighbor search can be found in [10].

The purpose of the filtering step (4th in Fig. 1) is to remove false matches that inevitably arise in all methods. A common procedure is to remove spatially close matches that appear because of correlation between spatially close regions. Another heuristic is skipping low-variance areas, such as skies, building facades, water surfaces etc. The main contribution of this paper belongs to this phase, filtering pairs of regions based on the compatibility with the JPEG compression process (JPEG copy-move constraint).

Finally, a postprocessing step verifies the spatial coherence of shift vectors (or in general transform parameters) obtained from the candidates generated by matching and filtering, as a rule by a combination of ideas known in image processing as the Hough transform [12] and RANSAC [14]. The seminal paper [15] simply considered only pairs with a shift vector identical to a sufficient number of other pairs. Postprocessing can also eliminate objects smaller than a threshold.

The block-based techniques are time-consuming due to the large number of compared blocks and lose accuracy when the tampered areas are blurred, scaled, rotated or otherwise geometrically transformed. To address these problems, some authors, instead of working with blocks, applied keypoint-based techniques used in computer vision, where the task is to find point correspondence between two images of one object or the same scene. Huang et al. [16] came up with a CMFD method based on the SIFT features [20], Bo et al. [8] used SURF [5]. In general, the keypoint-based techniques are fast and can be easily used when the patches were deformed by a geometric transform. On the down-side, they do not work well for small objects and are less stable than block-based methods. Especially, since the keypoints are usually detected in regions with high entropy, these methods lose accuracy in the areas retouched by an indistinct texture [35].

Hybrid schemes combine advantages of block-based and keypoint-based approaches [2, 26, 19, 36, 18]. Postprocessing can also include segmentation to estimate the contours of the copied object [33].

This paper assumes the basic but very frequent scenario of simple copy-move in JPEG images, i. e. without geometric distortions, added noise or blurring. We analyze the influence of JPEG compression and propose an algorithm that uses the principles of JPEG compression to increase the sensitivity of detection.

### 3. JPEG compression

JPEG is undoubtedly the most widespread format for efficient storage of image data [23]. JPEG uses a lossy type of compression based on the quantization of discrete cosine transform (DCT) coefficients, where the two-dimensional

DCT is applied to small blocks of usually  $8 \times 8$  pixels. In this section we describe this process in detail and introduce the mathematical notation needed later.

For grayscale images, the lossy part of JPEG compression can be expressed [34, 31] as

$$y = [QCx], \quad (1)$$

where  $x$  is a vectorized original image,  $y$  the vector of integer coefficients stored in the JPEG file,  $Q$  and  $C$  are matrices, and the square brackets denote the operator of rounding to the nearest integer.  $C$  is the block-diagonal matrix of the block DCT made up of the square matrices of the two-dimensional DCT ( $64 \times 64$  matrix for each  $8 \times 8$  block).  $C$  is orthogonal, because all the DCT sub-matrices are orthogonal.  $Q$  is a diagonal matrix corresponding to the element-wise division by the coefficients from the quantization table stored in each JPEG file replicated for each block along diagonal (64 values for each  $8 \times 8$  block). We will denote the vector of these coefficients as  $q$ , i.e.  $Q = \text{diag}(1/q)$ . For color images, the image is first transformed into  $Y'C_B C_R$  space and individual channels are stored separately. The brightness  $Y'$  is treated as described above but the chrominance channels are often stored at smaller resolution, which complicates the degradation model. Formally,

$$y = [QCDx], \quad (2)$$

where  $D$  is a down-sampling matrix. As a rule,  $D$  returns the average value for every (non-overlapping) square of  $2 \times 2$  pixels but other dimensions of down-sampling windows, such as  $2 \times 1$  or  $1 \times 2$  are possible. The grayscale case (1) can be considered a special case of (2) with  $D = I$ , i.e. identity. JPEG decompression can be written as

$$\tilde{x} = \frac{1}{k} D^T C^T Q^{-1} y = \frac{1}{k} D^T C^T \text{diag}(q) y, \quad (3)$$

where  $k = 1/4$  for the  $2 \times 2$  down-sampling window,  $k = 1/2$  for  $2 \times 1$  and  $1 \times 2$ , and  $k = 1$  for grayscale (no down-sampling).

Given coefficients  $y$  stored in a JPEG file, the quantization constraint set is the set of images satisfying condition (2) in all color channels. This set is local in the sense that it is defined independently for each JPEG block. This locality is reflected in the structure of matrix  $QCD$ , which is block-diagonal with blocks of size  $64 \times 64$  for grayscale and  $64 \times 256$  for the chrominance channels with  $2 \times 2$  down-sampling.

#### 4. JPEG copy-move constraint

The main contribution of this paper is a procedure to verify the possibility of copy-move between two patches with known coordinates using the knowledge of JPEG compression process including the quantization table extracted from the input JPEG file. We assume that the object was only moved, i.e. there was no rotation or any other geometrical distortion, and that there was no noise

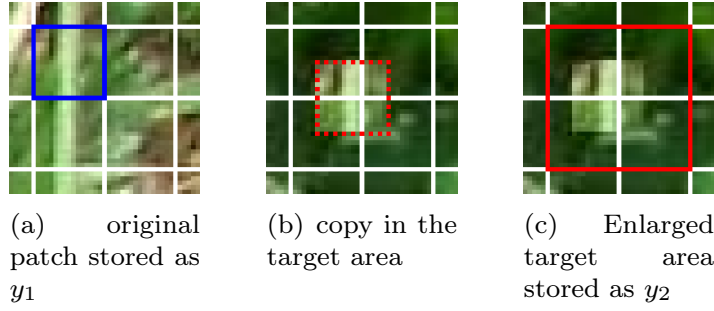


Figure 2: Alignment of original and target patches as used in the verification of the JPEG copy-move constraint.

added. The object could have been copied to an arbitrary part of the image and after moving to the target area, the object could have been retouched by blurring boundaries to mask the transition to the surrounding area. The main requirement is that the original object contains at least one JPEG block, i.e. a rectangular area of size at least  $8 \times 8$  pixels aligned with the JPEG grid (see Fig. 2(a)). This is guaranteed for objects of size  $15 \times 15$  pixels and larger. The corresponding target area (Fig. 2(b)) must not be retouched. If we use also the chromatic channels, the necessary patch size increases to  $16 \times 16$  pixels.

For our purpose, we assume that the detection algorithm works with patches of size being a multiple of the size of JPEG blocks (typically  $8 \times 8$  for grayscale and  $16 \times 16$  if we consider also the chrominance channels). In addition, we assume that the source patch is aligned with the JPEG grid (in Fig. 2(a) the patch consists of only one JPEG block). In the target area, we consider a larger patch, the smallest patch containing the cloned area aligned with JPEG blocks, see Fig. 2(c). Denoting the vectorized pixel values before compression in the enlarged target area as  $x$ , the values of the original source patch (Fig. 2(a)) can be expressed as  $Mx$ , where  $M$  is the matrix selecting the pixels corresponding to the source area in Fig. 2(b) (selection matrix). Equation (2) implies that the set of possible target areas before JPEG compression satisfy

$$QCDx \in \left\langle y_2 - \frac{1}{2}, y_2 + \frac{1}{2} \right\rangle, \quad (4)$$

where  $y_2$  is the vector of integer coefficients stored in the JPEG file that corresponds to the target patch. The interval  $\left\langle y_2 - \frac{1}{2}, y_2 + \frac{1}{2} \right\rangle$  is the multi-dimensional interval of numbers rounding to the nearest integers in  $y_2$ . The interval is left-closed (angle bracket) and right-open (round bracket) as usual when rounding. In the original area containing the same (source) patch  $Mx$  similarly

$$QCDMx \in \left\langle y_1 - \frac{1}{2}, y_1 + \frac{1}{2} \right\rangle. \quad (5)$$

Sets (4) and (5) are convex. We can see that the necessary condition for the possibility of copy-move between source and target areas is equivalent to the existence of a point in the intersection of sets (4) and (5). We call this condition

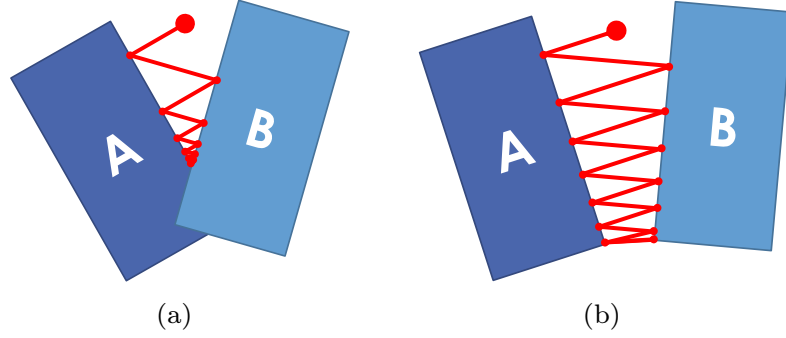


Figure 3: Finding the intersection of two sets by alternating projection, where  $A \cap B \neq \emptyset$  (a) or  $A \cap B = \emptyset$  (b). We use a faster variant of this algorithm to verify the proposed JPEG copy-move constraint.

the *JPEG copy-move constraint* or shortly *JPEG constraint*. Unfortunately, it is probably impossible to verify the existence of this intersection directly. Nevertheless, in the following section, we derive an efficient algorithm that verifies the JPEG constraint iteratively.

## 5. Algorithm

The simplest way to find a point in the intersection of arbitrary two convex sets  $A$  and  $B$  in a vector space is alternating projection (see Fig. 3). If the intersection is not empty, this procedure provably converges to a point in the intersection. Otherwise, the distance between two consecutive projections converges to the distance between  $A$  and  $B$ . This method is known as the projection on convex sets (POCS) [4]. Unfortunately, POCS converges relatively slowly (in our case hundreds of iterations), which makes it unsuitable for our purpose.

The convergence can be significantly speeded up by the Douglas-Rachford splitting [11], which is a special case of the alternating direction method of multipliers (ADMM) [13]. If exists, the intersection  $A \cap B$  can be found by iterating the following three steps

$$x \leftarrow P_A(a + d), \quad (6)$$

$$a \leftarrow P_B(x - d), \quad (7)$$

$$d \leftarrow d - (x - a), \quad (8)$$

where  $P_A$  and  $P_B$  are projections on sets  $A$  and  $B$ , and  $a$  and  $d$  are auxiliary variables of the same size as  $x$ . By the projection on a set we mean the point in the set closest to the projected point in the sense of  $l_2$  norm. Variable  $x$  is in general a point in set  $A$ , which will be in our algorithm the same  $x$  defined in the previous section. Variable  $a$  is initialized in a starting point  $x_0$ ,  $d$  is initially zero. Iterations are stopped, when the  $l_2$  norm of  $x - a$  is smaller than a threshold  $\varepsilon$  or the number of iterations exceeds a limit. The latter implies that  $A \cap B$  is empty. Note that both POCS and ADMM require sets  $A$  and  $B$

to be closed. For this reason, instead of intervals in (4) and (5), we work with closed intervals  $\langle y - \frac{1}{2}, y + \frac{1}{2} - \delta \rangle$ , where  $\delta$  is a sufficiently small constant.

To make the algorithm efficient, we need also the projections in (6) and (7) to be fast. Since the set (4) is of the same form as (5), it is sufficient to show the efficient projection for the latter. Indeed, projection (5) can be expressed as

$$P_{QCDMx \in \langle y - \frac{1}{2}, y + \frac{1}{2} - \delta \rangle}(z) = z - \frac{1}{k} M^T D^T C^T \text{diag}(q) \cdot \left( QCDMz - P_{\langle y - \frac{1}{2}, y + \frac{1}{2} - \delta \rangle}(QCDMz) \right), \quad (9)$$

which is straightforward to compute in time proportional to the number of pixels in the patch. Transpose  $C^T$  is the inverse DCT,  $D^T$  replicates each value in a down-sampled image to the corresponding rectangle in the full size image and  $M^T$  keeps the pixels selected by  $M$  (corresponding to the source patch) intact and fills the rest of the pixels of the target area with zeros. Constant  $k = 1/mn$  is a down-sampling factor ( $1/4$  for  $2 \times 2$  down-sampling). The projection on set (4) is a special case with  $M = I$ , i.e. identity. Informal proof (using a lemma from [30]) is given in the appendix.

The algorithm is summarized as Alg. 1. We can see that (9) and therefore also the algorithm basically consists of JPEG compression and decompression operations (compare with (2) and (3)). The compression operations  $QCD$  and  $QCDM$  are computed only once in each projection, since its result can be reused. In our experiments the threshold in the stopping criterion was set to  $\varepsilon = 10^{-10}$  and the maximum number of iterations to 12. What is important, the number of iterations does not depend on the number of image pixels. In theory, it can slightly increase for larger patches, though. The initial estimate  $x_0$  can be set to the values in the target area.

The time needed to compute one projection is approximately the same as the time of one JPEG compression ( $QCD$ ) and decompression ( $\frac{1}{k} D^T C^T \text{diag}(q)$ ) of an image of the same size as the patch. As a result, since there are two projections in one iteration, one on a source patch and one on a target patch, and we use twelve iterations, the time of Alg. 1 corresponds to 12 grayscale JPEG compressions and decompressions of an image of the same size as the aligned target patch, i.e.  $16 \times 16$  pixels for the source patches of size  $8 \times 8$ , and 12 of an image of the size of the source patch.

## 6. Experiments

In this section, we present results of a series of experiments demonstrating the benefits of the proposed JPEG copy-move constraint. Experiments in sections 6.1, 6.2 and 6.3 statistically analyze the influence of the constraint on the number of falsely detected moves, irrespectively of other steps in the chain of operations in Fig. 1, i.e. independently of what particular method is used. Section 6.4 shows the performance of the constraint in a complete detection algorithm. For this purpose, we use the JPEG copy-move constraint with

---

**Algorithm 1** Verification of the JPEG copy-move constraint, i. e. the possibility of copy-move between two image patches.

---

1. Initialize  $a_0 = x_0$ ,  $d_0 = 0$
2. repeat
- 3.

$$x \leftarrow a + d - \frac{1}{k} D^T C^T \text{diag}(q) \cdot \left( QCD(a + d) - P_{\langle y_2 - \frac{1}{2}, y_2 + \frac{1}{2} - \delta \rangle} (QCD(a + d)) \right)$$

- 4.

$$a \leftarrow x - d - \frac{1}{k} M^T D^T C^T \text{diag}(q) \cdot \left( QCDM(x - d) - P_{\langle y_1 - \frac{1}{2}, y_1 + \frac{1}{2} - \delta \rangle} (QCDM(x - d)) \right)$$

5.  $d \leftarrow d - (x - a)$
  6. until the stopping criterion  $\|x - a\| \leq \varepsilon$  is satisfied or the number of iterations exceeds a limit
- 

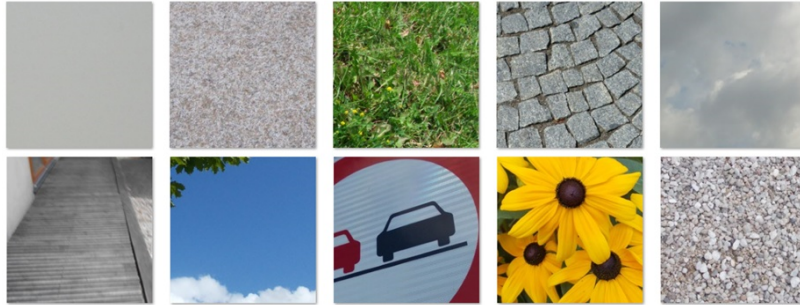


Figure 4: Images used to estimate the parameters of the algorithm and analyze the number of false matches.

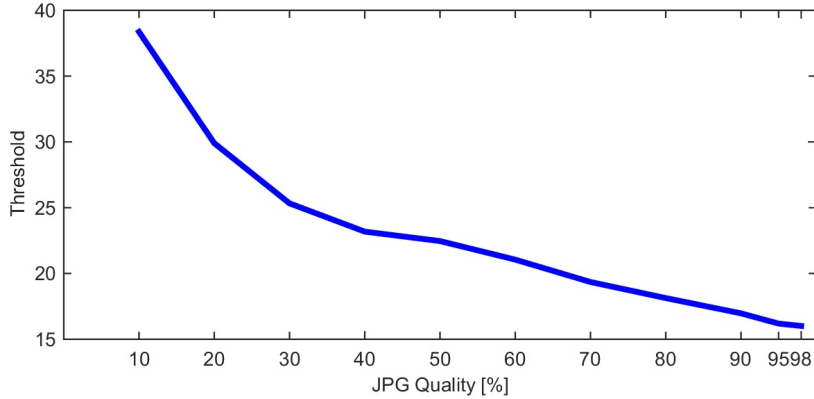


Figure 5: Maximum  $l_2$  distance per pixel of copy-moved patches at the sensitivity level 99.5%.

a simple patch-level matching, finally verifying the coherence of shift vectors. This approach is compared to two representatives of block-based techniques and one method based on keypoints, using three different datasets. Two datasets [29, 28] are standard benchmark datasets available online, the third was created by us to show performance on difficult cases with repeated patterns and indistinct texture.

### 6.1. Estimating the parameters of the algorithm

Purpose of the first experiment is to determine two parameters the algorithm requires, the detection threshold  $\varepsilon$  and the number of iterations the Alg. 1 requires to reliably verify the JPEG constraint. We use ten images of size  $1024 \times 1024$  pixels (Fig. 4), covering several types of textures. We select randomly a position of source patch aligned with the JPEG grid, copy the data, paste the patch to a random off-grid position and save this new picture to JPEG. The patches pasted on the positions aligned with the grid are not considered, because they keep exactly the same values. After reloading the image, we take both patches, compute their  $l_2$  distance per pixel and run our algorithm until convergence ( $\varepsilon = 10^{-10}$ ). Since these patches come from the same original patch, the algorithm always converges. For each iteration, we remember the norm of the residual  $\|x - a\|$ . We repeat this experiment 1000 times for each image, with the quality of JPEG set to values in  $\{10, 20, \dots, 90, 95, 98\}$ .

This experiment gives us two valuable outputs. The maximum value of the  $l_2$  distance gives us an estimate of the threshold necessary to select the pairs of patches that must be considered valid candidates for matching. However, for computational reasons, we select the threshold to detect only 99.5% of candidates. Otherwise, we would have to test an excessive number of candidate patches. The resulting threshold is shown in Fig. 5. The threshold decreases with increasing JPEG quality and stays surprisingly high (above 15) even for the best qualities.

The second output is the number of iterations the algorithm requires to tell whether the JPEG constraint is satisfied or not. Fig. 6 shows that the brightness component  $Y'$  never requires more than 12 iterations for any JPEG

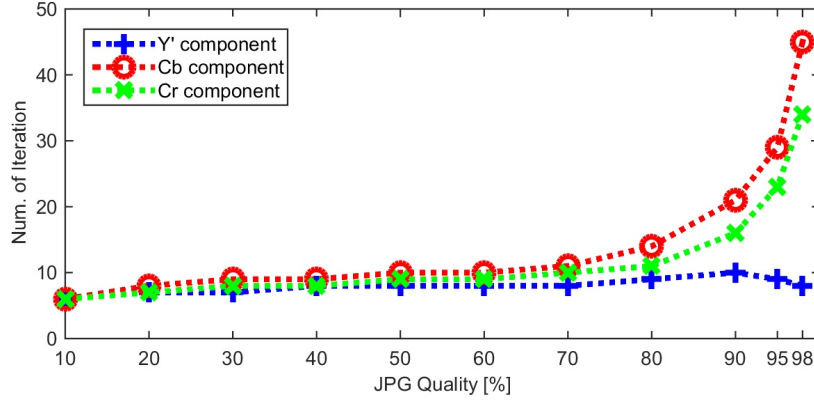


Figure 6: Number of iterations needed for convergence ( $\varepsilon = 10^{-10}$ ).

quality. This is in contrast to the chrominance components  $C_B$  and  $C_R$  that require up to 50 iterations for higher qualities. In practice, since the brightness channel carries more information than the chrominance channels, it is mostly sufficient to work only with brightness, which speeds up computation and does not considerably increase the number of false matches. In addition, we can use smaller patches of  $8 \times 8$  pixels, instead of  $16 \times 16$ , which is useful for detecting small objects. For this reason, in the rest of this section, all experiments work only with the brightness channel, i.e. in grayscale.

Our tests on the other image sets showed that the setting  $\varepsilon = 10^{-10}$  and 12 iterations is quite universal, which means that using the JPEG copy-move constraint does not require setting any additional parameters.

### 6.2. Reducing the number of false matches

The purpose of the second experiment is to demonstrate that the JPEG copy-move constraint decreases the number of false matches. The procedure is similar to what we did previously except that we do not clone the patches and assume that randomly taken pairs of patches should always be negatives. We take images from Fig. 4 and save them using the same set of compression qualities  $\{10, 20, \dots, 90, 95, 98\}$ . After reloading each image, we go patch by patch from the upper-left corner to the lower-right corner. For each patch, we find the most similar patch from the same image (in the  $l_2$  distance) and run our procedure for verifying the JPEG constraint.

If we use the thresholds computed in the first experiment using the  $l_2$  distance, the number of false positives is huge, over 60% even for qualities above 90 (solid line in Fig. 7). We cannot set the threshold lower, otherwise we could miss some small objects because of decreased sensitivity. Note that using the DCT features as in [15] would keep the percentage of false matches high and therefore, for our comparison, it would not help. The dotted line shows the percentage of false matches when we apply also the JPEG copy-move constraint. We can see that the number of false matches decreases. The effect is very strong at qualities higher than 90, where the percentage decreases to less than 1.5%. At quality 95, we get less than 0.04% and at 98 there are no false matches. Note

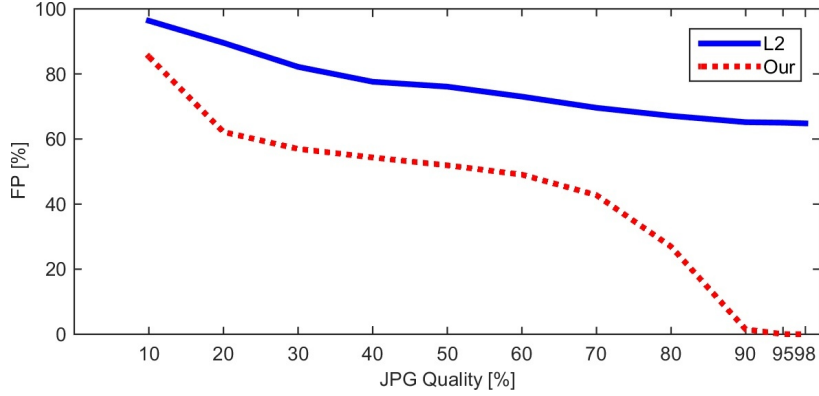


Figure 7: Percentage of false matches (false positives) at sensitivity level 99.5% (thresholds taken from Fig. 5, number of iterations 12)

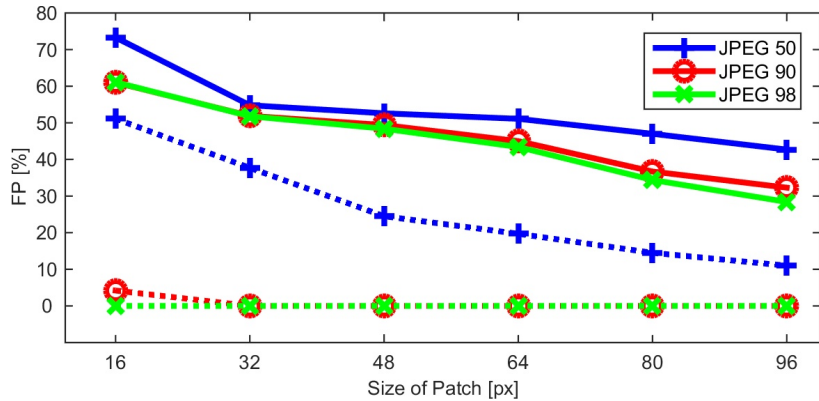


Figure 8: Percentage of false matches for three quality levels (50, 90 and 98) using the  $l_2$  distance (solid line) and the JPEG copy-move constraint (dotted) at sensitivity level 99.5%.

that the default JPEG qualities in the majority of modern photo editing tools and cameras are above 90 and therefore these high qualities are very common.

### 6.3. Influence of patch size

The third experiment (Fig. 8) shows the influence of the patch size. We considered patches of sizes  $16 \times 16$ ,  $32 \times 32$ ,  $48 \times 48$ ,  $64 \times 64$ ,  $80 \times 80$  and  $96 \times 96$  pixels. As the patch size increases, the number of false matches slightly drops. For example, for quality 50 and the JPEG constraint, the percentage of false matches drops from 50% to 10%. Without the constraint, this number remains as high as 30%.

### 6.4. Comparison on benchmark datasets

Previous experiments showed that the JPEG copy-move constraint decreases the number of false matches significantly. In this section, we show how this procedure improves performance of a complete detection algorithm. We use simple patch-level matching, followed by the JPEG copy-move constraint and verification of the coherence of shift vectors. For each JPEG block, we find the nearest patch in the  $l_2$  norm. Such pair of patches is considered a valid copy-move candidate if their  $l_2$  distance is smaller than a threshold determined by the

experiment described in section 6.1. The coherence is achieved by considering only the object with the highest number of detected shift vectors. We will see that even this simple variant gives impressive results. This approach is compared to three alternatives.

As a representative of block-based methods, we chose the method based on the quantization of the DCT coefficients from the seminal paper [15]. When used with  $8 \times 8$  patches needed to detect smaller objects, the basic version of the algorithm, as presented in the paper, gives relatively weak results. The main reason is that the algorithm uses lexicographic sorting and therefore misses many potential candidates. On the other hand, it is quite easy to tune up the method to give much better results by increasing the level of quantization and setting a suitable threshold to the maximum allowed distance of quantized DCT coefficients to be considered copy-move candidates. To distinguish these two versions, we denote them in our comparison as DCT-basic and DCT-tuned. To verify the coherence of shift vectors, both DCT-based variants we implemented use the same rule as the algorithm with the JPEG constraint. As a representative of methods based on keypoints, we use the SIFT-based algorithm [1], available online.

To simplify comparison to other methods in literature, we use two standard image datasets [29, 28]. The performance of other methods on the same datasets is analyzed in [9, 28]. We also created a third image set containing difficult cases with repeated patterns, objects masked by an indistinct texture (skies, building facades etc.) and small objects. All three datasets (DS1, DS2 and DS3), contain in total 81 images:

- **DS1:** 23 images by Silva et al. [28],
- **DS2:** 40 images, CoMoFoD dataset [29],
- **DS3:** 18 images created by our research group, representing the difficult cases.

To evaluate the algorithms, we use two different metrics. In both cases, we compare with the ground truth represented, for each image, by a matrix of ones on the tampered pixels and zeroes elsewhere.

- *Image level:* An algorithm is successful if it finds at least a portion of the tampered area. The algorithm fails if it does not find any modified pixel.
- *Patch level:* We divide images to non-overlapping patches of  $8 \times 8$  pixels. A patch is considered correctly marked as tampered if there is an overlap with a pixel in the ground truth. We count the number of correctly found tampered patches (true positives, TPs) and the number of patches labeled as tampered with no overlap with the ground true (false positives, FPs).

For space reasons, we show results on JPEG images compressed only with quality factor 95. As Fig. 7 suggests, we can expect similar behavior for qualities above around 90.

Quantitative results for all the datasets as well as for the union of all the datasets are given in Tables 1-4. The second column shows the number of images, where the algorithm found at least a part of the tampered area. The same number, as a percentage of all images in the dataset, is given in the third column (success rate). The forth and sixth columns show the numbers of patch level TPs and FPs. We also compute two standard statistical quantities, sensitivity (abbreviated as sens. in tables 1-4) and false discovery rate (FDR, one minus precision). They are given in columns five and seven. The number of patches in ground truth is defined as the number of patches containing at least one modified pixel. Three examples in Figs. 9, 10, and 11 illustrate typical outputs of the tested approaches.

At the image level, the SIFT-based method achieves a decent success rate of 91% on DB1 but often fails on DB2 and DB3 (65% and 33%). Moreover, this method produces a large number of false matches at the patch level, even for relatively easy examples like those in Figs. 9(e) and 10(e). The Fig. 11(e) illustrates how repeated motives can cause a complete failure of this algorithm, a problem mentioned already in the original paper [1].

The results in tables 1-4 and examples in Figs. 9(c) and 10(c) show that the basic version of the DCT-based algorithm [15] misses most true matches. Using the modified version of [15] described above (DCT-tuned), we achieved an overall success rate over 92% at the image level and sensitivity over 82% at the patch level, see the Tab. 4. As the weakest point of this approach, we identified the pictures with large areas of weak texture. This is demonstrated on two examples in Figs. 12 and 13, where the algorithm produces a large number of false matches in the homogeneous background, which cannot be distinguished from the true patches containing a relatively small object. We observed analogous behavior in many pictures with for example the skies, building facades or road surfaces.

The proposed JPEG copy-move constraint improves results significantly even when used with the simple  $l_2$  distance between patches. It achieves the overall success rate of 100% at the image level and sensitivity 89.91 % at the patch level. The main advantage of the JPEG constraint is that it increases precision and sensitivity by reducing the number of false positives (1.55%) even in extremely homogeneous areas like in Figs. 12 and 13. As a result, for larger objects, we get very compact results even without any additional postprocessing, see Figs. 9, 10 and 11. Note that several missing pixels in Figs. 9 and 10 are caused by considering, for speed reasons, only the closest patch. Using a more elaborate nearest neighbor search algorithm would easily remove them in these examples. They could be also quickly removed by verifying the JPEG constraint for all "holes" in the detected objects.

The computation of JPEG copy-move constraint comes at the cost of additional time, depending on how many candidate pairs we consider. In the simplified algorithm used in our experiments that work with only the closest patch to each 8x8 block on the JPEG grid, the estimate from the end of section 5 gives a rough upper bound of  $4 * 12 + 12 = 60$  JPEG compressions and decompressions of the analyzed image in the worst case, when all the pairs we found are close enough to be considered valid candidates, for example on an image

Method	Image level		Patch level TPs		Patch level FPs	
	#	%	#	sens. [%]	#	FDR [%]
SIFT	21	91.30	9127	66.89	1641	15.24
DCT - basic	16	69.57	2697	19.77	1855	40.75
DCT - tuned	23	100.00	11644	85.34	89	0.76
<b>OUR</b>	<b>23</b>	<b>100.00</b>	<b>12432</b>	<b>91.11</b>	<b>56</b>	<b>0.45</b>

Table 1: Comparison of methods on dataset DS1, 23 images, 13645 patches in ground-truth.

Method	Image level		Patch level TPs		Patch level FPs	
	#	%	#	sens. [%]	#	FDR [%]
SIFT	26	65.00	6057	65.21	9196	60.29
DCT - basic	16	40.00	532	5.73	1130	67.99
DCT - tuned	36	90.00	6601	71.07	595	8.27
<b>OUR</b>	<b>40</b>	<b>100.00</b>	<b>8055</b>	<b>86.72</b>	<b>127</b>	<b>1.55</b>

Table 2: Comparison of methods on dataset DS2, 40 images, 9288 patches in ground-truth.

of clear skies. For example, on our PC with Intel Core i5 CPU (3.4GHz) one JPEG compression/decompression of an image of size  $512 \times 512$  (DS2) using the Matlab JPEG toolbox takes about 1/40s, i.e.  $2 * 60/40 \sim 3s$  in total.

## 7. Conclusion

In this paper, we propose a method to improve reliability of detecting copy-move forgery in JPEG images. For this purpose, we introduce the JPEG copy-move constraint that can be used to filter out false matches of candidate patches in almost any existing detection algorithm. We show an efficient algorithm to verify this constraint. Since the constraint is exact, i.e. gives no false negatives, results are always better than without this constraint. For color images, the algorithm is derived for both the brightness and chrominance channels. However, since the brightness channel carries more information than chrominance channels, it is often sufficient to work only with brightness, which speeds up computation, allows for smaller patches ( $8 \times 8$  instead of  $16 \times 16$ ) and does

Method	Image level		Patch level TPs		Patch level FPs	
	#	%	#	sens. [%]	#	FDR [%]
SIFT	6	33.33	4845	49.19	4793	49.73
DCT - basic	10	55.56	2353	23.89	862	26.81
DCT - tuned	16	88.89	8672	88.04	890	9.31
<b>OUR</b>	<b>18</b>	<b>100.00</b>	<b>8987</b>	<b>91.24</b>	<b>119</b>	<b>1.31</b>

Table 3: Comparison of methods on dataset DS3 (difficult cases), 18 images, 9850 patches in ground-truth.

Method	Image level		Patch level TPs		Patch level FPs	
	#	%	#	sens. [%]	#	FDR [%]
SIFT	53	65.43	20029	61.10	15630	43.83
DCT - basic	42	51.85	5582	17.03	3847	40.80
DCT - tuned	75	92.59	26917	82.11	1574	5.52
<b>OUR</b>	<b>81</b>	<b>100.00</b>	<b>31301</b>	<b>89.91</b>	<b>302</b>	<b>1.01</b>

Table 4: Comparison of methods on full dataset (DS1+DS2+DS3), 81 images, 32783 patches in ground-truth.

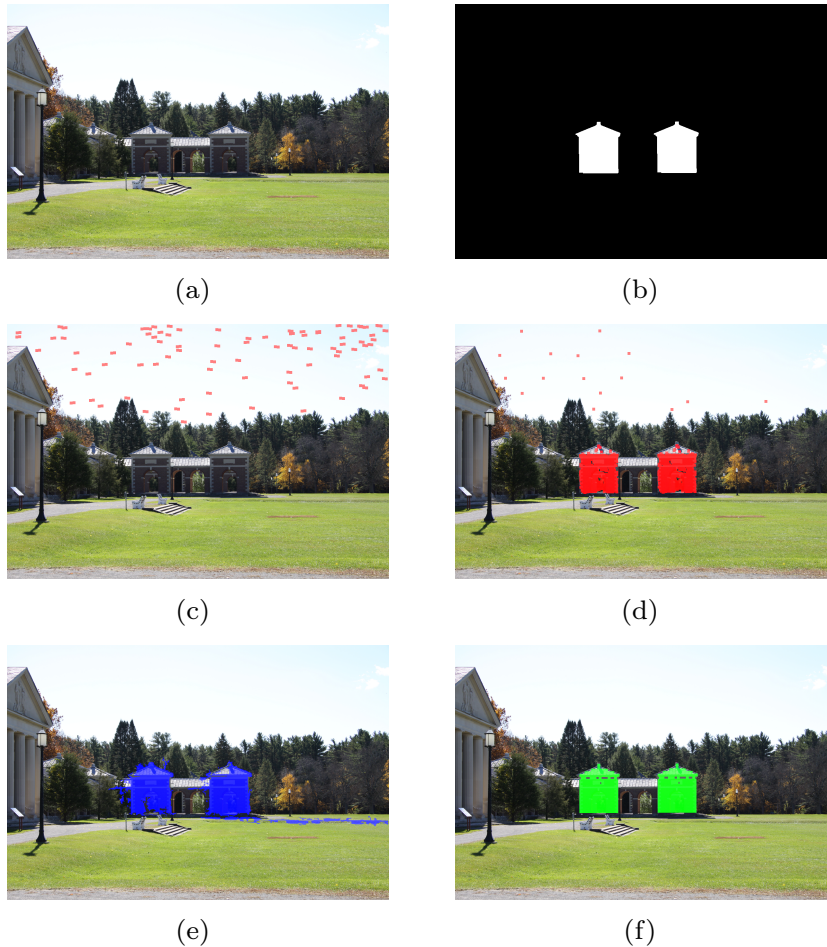


Figure 9: Example of detection. Figure (a) is an original image from DS1, (b) ground truth, (c) result of DCT-basic, (d) DCT-tuned, (e) SIFT, and (f) proposed algorithm with JPEG copy-move constraint.

not considerably increase the number of false matches.

Experiments show that the JPEG copy-move constraint is very strong for JPEG quality factors above 90, where it almost completely eliminates false matches. However, the reduction of false positives is considerable already from qualities around 70. The JPEG constraint is useful especially for small objects and areas with indistinct texture, where using the coherence of shift vectors in the postprocessing phase is not sufficient. This phenomenon is less pronounced

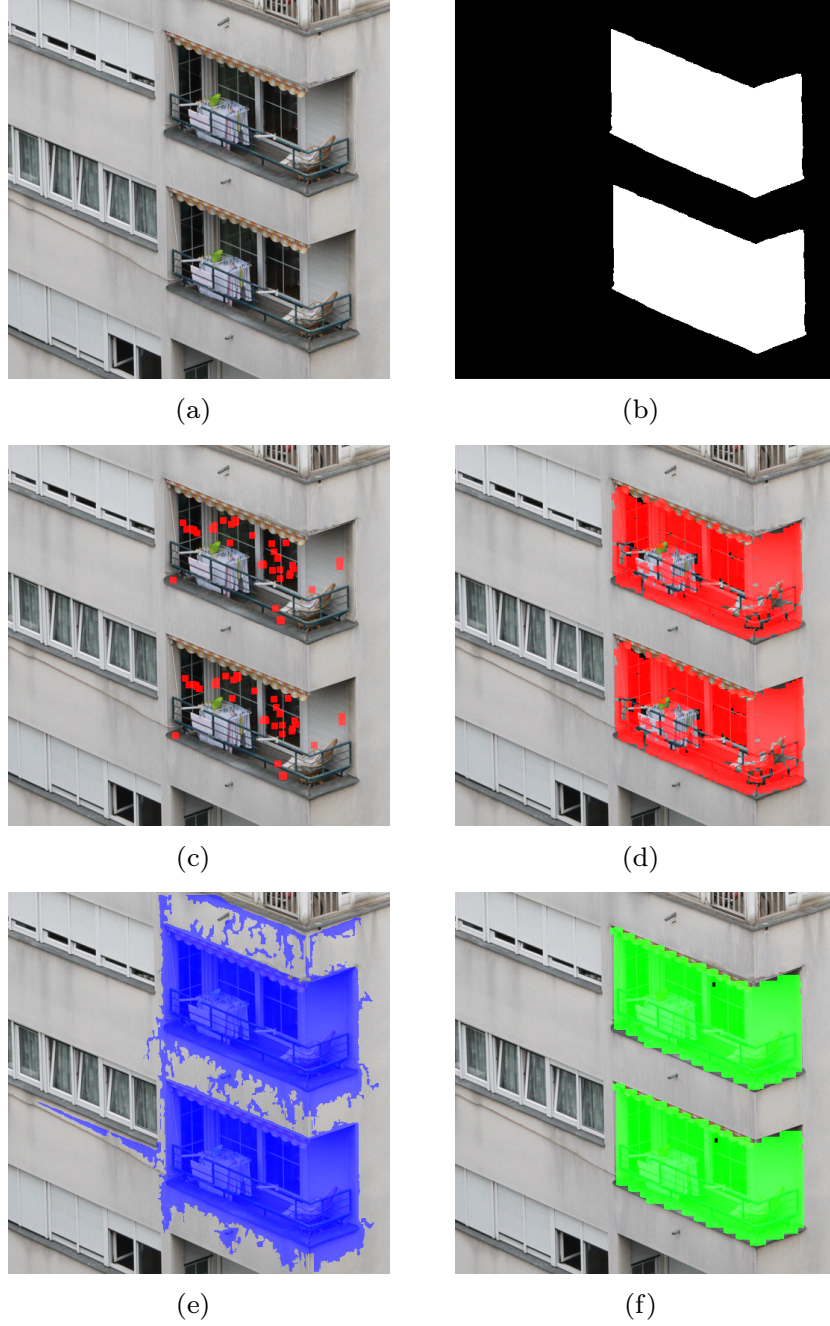


Figure 10: Example of detection. Figure (a) is an original image from DS2, (b) ground truth, (c) result of DCT-basic, (d) DCT-tuned, (e) SIFT, and (f) proposed algorithm with JPEG copy-move constraint.

for larger objects but the statistics in the experimental section show that even there the JPEG constraint helps significantly. The JPEG constraint can also distinguish the copy-move forgery from naturally repeating patterns, which is very hard for block-based methods and probably impossible for the methods based on keypoints.

To demonstrate the influence of the JPEG copy-move constraint, we used a relatively simple algorithm. We can expect that using the constraint with a

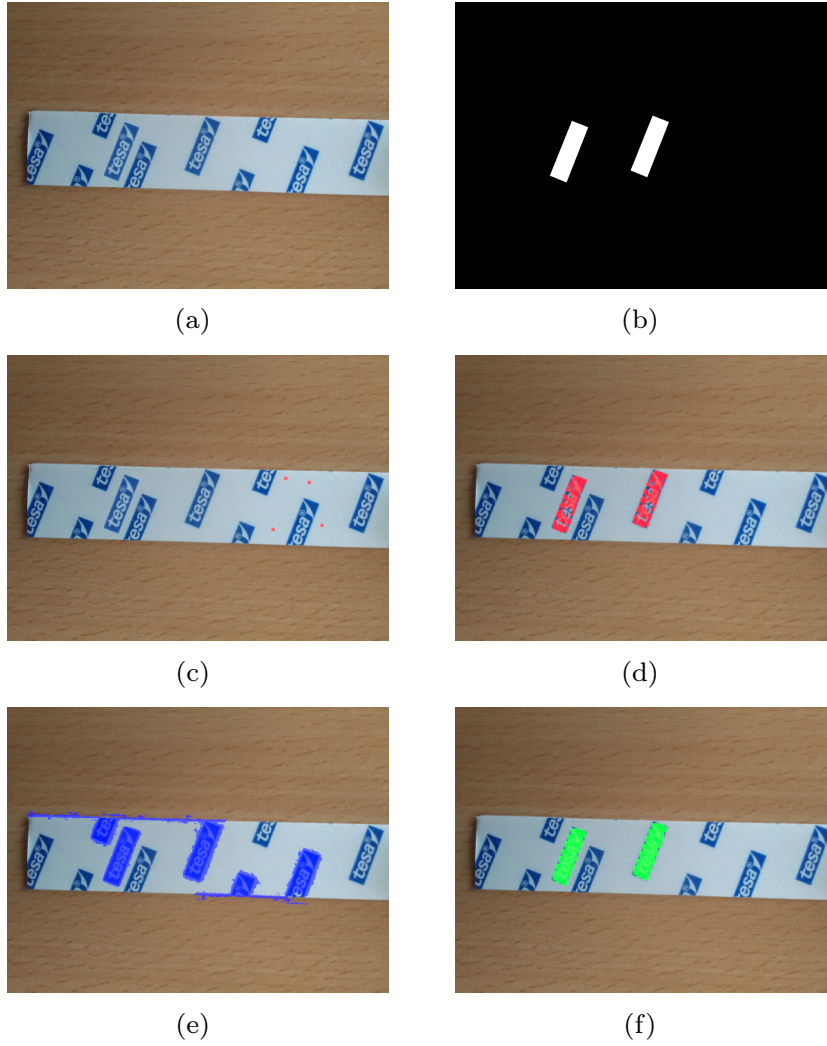


Figure 11: Example of detection. Figure (a) is an original image from DS3, (b) ground truth, (c) result of DCT-basic, (d) DCT-tuned, (e) SIFT, and (f) proposed algorithm with JPEG copy-move constraint.

more elaborated algorithm the results could be even better.

The proposed approach cannot be directly used with geometrical transforms like scale change or rotation. One could imagine an extension to these operations but results probably would not be worth additional complexity of the algorithm. In our opinion, improved detection justifies this restriction, though. In addition, the scenario of pure shift is probably frequent enough to be considered separately. Another limitation of the JPEG constraint is that it requires the moved object to contain at least one JPEG block. On the other hand, smaller objects are probably too hard also for all other methods.

## Acknowledgment

This research was partially funded by the Czech Science Foundation, grant GA16-13830S.

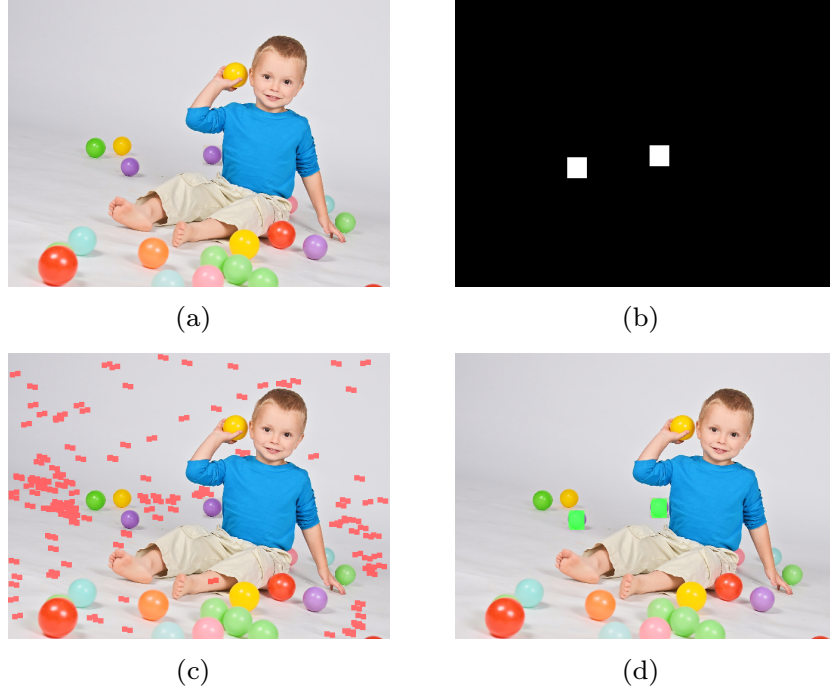


Figure 12: Reducing false matches. Figure (a) is an original image from DS3, (b) ground truth, (c) result of DCT-tuned, (d) proposed algorithm with JPEG copy-move constraint. The SIFT method failed completely.

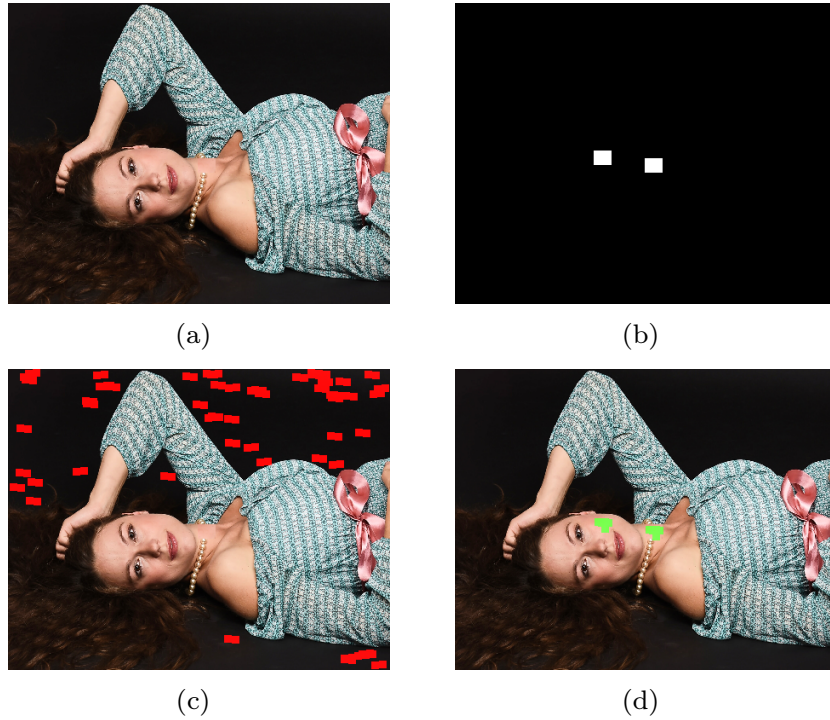


Figure 13: Reducing false matches. Figure (a) is an original image from DS3, (b) ground truth, (c) result of DCT-tuned, (d) proposed algorithm with JPEG copy-move constraint. The SIFT method failed completely (found nothing).

## Appendix

### Proposition 1. Projection

$$P_{QCDMx \in \langle y - \frac{1}{2}, y + \frac{1}{2} - \epsilon \rangle}(z) = z - \frac{1}{k} M^T D^T C^T \text{diag}(q) \cdot \left( QCDM(z) - P_{\langle y - \frac{1}{2}, y + \frac{1}{2} - \epsilon \rangle}(QCDMz) \right). \quad (.1)$$

*Proof.* The formula for the projection is a consequence of Lemma 2. To be applicable, we must show that  $AA^T$ , where  $A = QCDM$ , is diagonal. First,  $M^T$  takes a source patch and expands it to the size of the target patch by filling with zeros and  $M$  selects back the pixels of the source patch, therefore  $MM^T = I$ . Second, let us assume that  $D$  is the averaging over non-overlapping windows of  $m \times n$  pixels. Then  $D^T$  replicates each pixel of a down-sampled image to a  $m \times n$  window multiplied by a factor  $k = 1/mn$ . Since  $D$  only averages the values in the window, we are getting  $DD^T = I/mn = kI$ . Finally, since  $C$  is orthogonal, i.e.  $CC^T = I$ , we are getting  $AA^T = QCDMM^T D^T C^T Q^T = QCDD^T C^T Q^T = kQCC^T Q^T/mn = kQQ^T$ , which is diagonal. Finally,  $A^T(AA^T)^{-1} = M^T D^T C^T Q^T \frac{1}{k} \text{diag}(q^2) = \frac{1}{k} M^T D^T C^T \text{diag}(q)$ .  $\square$

**Lemma 2.** Projection  $P_{Ax \in \langle b_1, b_2 \rangle}(z) = \arg \min_x \|x - z\|$ , s.t.  $Ax \in \langle b_1, b_2 \rangle$ ,  $b_1 \leq b_2$ ,  $A \in R^{m \times n}$ ,  $m \leq n$  full rank,  $AA^T$  diagonal, can be computed as  $P_{Ax \in \langle b_1, b_2 \rangle}(z) = z - A^T(AA^T)^{-1}(Az - P_{\langle b_1, b_2 \rangle}(Az))$ , where  $P_{\langle b_1, b_2 \rangle}(y) = \min(\max(b_1, y), b_2)$ .

The proof of this lemma can be found in the appendix of [30].

- 
- [1] Amerini, I., Ballan, L., Caldelli, R., Del Bimbo, A., Serra, G., Sept 2011. A sift-based forensic method for copy-move attack detection and transformation recovery. IEEE Transactions on Information Forensics and Security 6 (3), 1099–1110.
  - [2] Ardizzone, E., Bruno, A., Mazzola, G., Oct 2015. Copy-move forgery detection by matching triangles of keypoints. IEEE Transactions on Information Forensics and Security 10 (10), 2084–2094.
  - [3] Bashar, M. K., Noda, K., Ohnishi, N., Mori, K., 2010. Exploring duplicated regions in natural images. IEEE Transactions on Image Processing PP (99), 1–1.
  - [4] Bauschke, H. H., Borwein, J. M., 1996. On projection algorithms for solving convex feasibility problems. SIAM review 38 (3), 367–426.
  - [5] Bay, H., Ess, A., Tuytelaars, T., Gool, L. V., 2008. Speeded-up robust features (surf). Computer Vision and Image Understanding 110 (3), 346 – 359, similarity Matching in Computer Vision and Multimedia.

- [6] Bayram, S., Sencar, T. H., Memon, N., 2009. An efficient and robust method for detecting copy-move forgery. In: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 1053–1056.
- [7] Beis, J. S., Lowe, D. G., 1997. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In: Computer Vision and Pattern Recognition, 1997. Proceedings. IEEE, pp. 1000–1006.
- [8] Bo, X., Junwen, W., Guangjie, L., Yuewei, D., 2010. Image copy-move forgery detection based on surf. In: 2010 International Conference on Multimedia Information Networking and Security. pp. 889–892.
- [9] Christlein, V., Riess, C., Jordan, J., Riess, C., Angelopoulou, E., Dec 2012. An evaluation of popular copy-move forgery detection approaches. IEEE Transactions on Information Forensics and Security 7 (6), 1841–1854.
- [10] Cozzolino, D., Poggi, G., Verdoliva, L., 2015. Efficient dense-field copy-move forgery detection. IEEE Transactions on Information Forensics and Security 10 (11), 2284–2297.
- [11] Douglas, J., Rachford, H. H., 1956. On the numerical solution of heat conduction problems in two and three space variables. Transactions of the American mathematical Society, 421–439.
- [12] Duda, R. O., Hart, P. E., 1972. Use of the hough transformation to detect lines and curves in pictures. Communications of the ACM 15 (1), 11–15.
- [13] Eckstein, J., Bertsekas, D. P., June 1992. On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. 55, 293–318.
- [14] Fischler, M. A., Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24 (6), 381–395.
- [15] Fridrich, J., Soukal, D., Lukas, J., Aug. 2003. Detection of copy move forgery in digital images. In: Digital Forensic Research Workshop.
- [16] Huang, H., Guo, W., Zhang, Y., Dec 2008. Detection of copy-move forgery in digital images using sift algorithm. In: Computational Intelligence and Industrial Application, 2008. PACIIA '08. Pacific-Asia Workshop on. Vol. 2. pp. 272–276.
- [17] Huang, Y., Lu, W., Sun, W., Long, D., 2011. Improved dct-based detection of copy-move forgery in images. Forensic Science International 206 (1-3), 178–184.
- [18] Li, H., Luo, W., Qiu, X., Huang, J., May 2017. Image forgery localization via integrating tampering possibility maps. IEEE Transactions on Information Forensics and Security 12 (5), 1240–1252.

- [19] Li, J., Li, X., Yang, B., Sun, X., 2015. Segmentation-based image copy-move forgery detection scheme. *IEEE Transactions on Information Forensics and Security* 10 (3), 507–518.
- [20] Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60 (2), 91–110.
- [21] Mahdian, B., Saic, S., 2007. Detection of copy-move forgery using a method based on blur moment invariants. *Forensic Science International* 171 (2-3), 180–189.
- [22] Muja, M., Lowe, D. G., 2009. Fast approximate nearest neighbors with automatic algorithm configuration. *VISAPP* (1) 2, 331–340.
- [23] Pennebaker, W. B., Mitchell, J. L., 1992. *JPEG Still Image Data Compression Standard*, 1st Edition. Kluwer Academic Publishers, Norwell, MA, USA.
- [24] Piva, A., 2013. An overview on image forensics. *ISRN Signal Processing* 2013.
- [25] Popescu, A. C., Farid, H., 2004. Exposing digital forgeries by detecting duplicated image regions. Tech. Rep. TR2004-515, Department of Computer Science, Dartmouth College.  
URL [www.cs.dartmouth.edu/farid/publications/tr04.html](http://www.cs.dartmouth.edu/farid/publications/tr04.html)
- [26] Pun, C.-M., Yuan, X.-C., Bi, X.-L., Aug 2015. Image forgery detection using adaptive oversegmentation and feature point matching. *IEEE Transactions on Information Forensics and Security* 10 (8), 1705–1716.
- [27] Ryu, S.-J., Lee, M.-J., Lee, H.-K., 2010. Information Hiding: 12th International Conference, IH 2010, Calgary, AB, Canada, June 28-30, 2010, Revised Selected Papers. Springer Berlin Heidelberg, Berlin, Heidelberg, Ch. Detection of Copy-Rotate-Move Forgery Using Zernike Moments, pp. 51–65.
- [28] Silva, E., Carvalho, T., Ferreira, A., Rocha, A., 2015. Going deeper into copy-move forgery detection: Exploring image telltales via multi-scale analysis and voting processes. *Journal of Visual Communication and Image Representation* 29, 16 – 32.
- [29] Tralic, D., Zupancic, I., Grgic, S., Grgic, M., Sept 2013. CoMoFoD-New database for copy-move forgery detection. In: *Proceedings ELMAR-2013*. pp. 49–54.
- [30] Šorel, M., Bartoš, M., 2016. Efficient jpeg decompression by the alternating direction method of multipliers. In: *International Conference on Pattern Recognition, ICPR’16*.

- [31] Šorel, M., Bartoš, M., Jan 2017. Fast bayesian jpeg decompression and denoising with tight frame priors. *IEEE Transactions on Image Processing* 26 (1), 490–501.
  - [32] Wang, J., Liu, G., Li, H., Dai, Y., Wang, Z., 2009. Detection of image region duplication forgery using model with circle block. In: 2009 International Conference on Multimedia Information Networking and Security. Vol. 1. pp. 25–29.
  - [33] Wang, W., Dong, J., Tan, T., 2014. Exploring dct coefficient quantization effects for local tampering detection. *IEEE Transactions on Information Forensics and Security* 9 (10), 1653–1666.
  - [34] Yang, Y., Galatsanos, N. P., Katsaggelos, A. K., 1995. Projection-based spatially adaptive reconstruction of block-transform compressed images. *Image Processing, IEEE Transactions on* 4 (7), 896–908.
  - [35] Yu, L., Han, Q., Niu, X., 2016. Feature point-based copy-move forgery detection: covering the non-textured areas. *Multimedia Tools and Applications* 75 (2), 1159–1176.
  - [36] Zandi, M., Mahmoudi-Aznaveh, A., Talebpour, A., Nov 2016. Iterative copy-move forgery detection based on a new interest point detector. *IEEE Transactions on Information Forensics and Security* 11 (11), 2499–2512.
-



# Automatic blood detection in capsule endoscopy video

Adam Novozámský,<sup>a</sup> Jan Flusser,<sup>a,\*</sup> Ilya Tachecí,<sup>b</sup> Lukáš Sulík,<sup>c</sup> Jan Bureš,<sup>b</sup> and Ondřej Krejcar<sup>c</sup>

<sup>a</sup>Institute of Information Theory and Automation, The Czech Academy of Sciences, Pod vodárenskou věží 4, 182 08 Prague 8, Czech Republic

<sup>b</sup>Charles University, Second Department of Internal Medicine-Gastroenterology, Faculty of Medicine and University Hospital, Sokolská 581, Hradec Králové 500 05, Czech Republic

<sup>c</sup>University of Hradec Králové, Center for Basic and Applied Research, Faculty of Informatics and Management, Hradecká 1249, Hradec Králové 500 03, Czech Republic

**Abstract.** We propose two automatic methods for detecting bleeding in wireless capsule endoscopy videos of the small intestine. The first one uses solely the color information, whereas the second one incorporates the assumptions about the blood spot shape and size. The original idea is namely the definition of a new color space that provides good separability of blood pixels and intestinal wall. Both methods can be applied either individually or their results can be fused together for the final decision. We evaluate their individual performance and various fusion rules on real data, manually annotated by an endoscopist. © 2016 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JBO.21.12.126007](https://doi.org/10.1117/1.JBO.21.12.126007)]

**Keywords:** wireless capsule endoscopy; bleeding detection; color transformation; morphology; decision fusion.

Paper 160460R received Jul. 1, 2016; accepted for publication Nov. 17, 2016; published online Dec. 9, 2016.

## 1 Introduction

Wireless capsule video endoscopy (WCE) is a noninvasive diagnostic tool for small bowel investigation that has been used in clinical praxis since 2001. WCE is very considerate of the patient. The patient swallows a cylindrical plastic capsule of a size about  $10 \times 25$  mm (depending on the manufacturer), which contains a digital video camera, LED light source, signal transmitter, and battery. The capsule travels through the gastrointestinal tract by peristaltic contractions, captures the images, and wirelessly transmits them in real time to an external console worn by the patient. The images are recorded and stored in the console memory and can be uploaded to a computer for a visual inspection or automatic analysis immediately after the monitoring has been completed. Current capsules take frames at a rate between two and six per second on average, which results in many thousands of images (typically up to 60,000) and  $>10$  h of video per patient. (Fortunately, current batteries are powerful enough to supply the light source, camera, and transmitter all the time.) The primary use of the capsule endoscopy is to examine areas of the small intestine that are difficult to display by other types of endoscopy. WCE has been successfully applied in the detection of small bowel bleeding sources, Crohns disease, complications of coeliac disease, small bowel tumors, and nonsteroidal anti-inflammatory drug induced enteropathy.

The main obstacle in a routine usage of WCE is that the visual evaluation of the video is very time consuming. It is supposed to be done by a trained endoscopist. Even if the software provided by the capsule producers offers certain supporting tools to simplify and speed up the process, it still takes at least 1 h of full concentration of the evaluator. Taking into account that the pathology can be visible only on very few frames and, hence, can be missed easily, the importance of a human factor is apparent.

The goal of this paper is to propose a technique for the detection of frames that are suspected to have the presence of bleeding. We do not aim to develop a fully automatic tool for bleeding detection that would replace the doctor. Rather, the method should preprocess the video, identify and export suspected frames, and prepare them for visual inspection, whereas the other frames are skipped and not sent to the inspection at all. This may significantly reduce the evaluation time while the final decision is still left to the endoscopist. This intended goal predetermines the required properties of the method. It should be fast and should provide a high true-positive (TP) rate while keeping the false-positive (FP) rate reasonably low.

In Sec. 2, we present a short survey of relevant literature. Section 3 describes the proposed algorithms. In Sec. 4, we present the WCE technical parameters and implementation details of the methods. Section 5 contains experimental evaluation on real data.

## 2 Literature Survey

The first research articles on WCE video analysis appeared soon after the WCE had been introduced into clinical praxis. Since the bleeding detection is a very frequent requirement and appears to be easily achievable (which is, however, not generally true as we can see from the literature survey), the largest group of papers has dealt with this challenge. The main problem here is that the blood spots and traces do not have any typical texture and shape, and the blood color may vary widely from light red through dark red to brown, which makes the blood difficult to distinguish from other objects and from the intestinal wall. This diversity in color depends on the types of disease, the bleeding time,<sup>1</sup> the position of the capsule, and the surrounding conditions.<sup>2</sup> If the parameters have been trained on a certain data collection, the color-based methods perform well on this set, but their generalization to other patients or even to another capsule

\*Address all correspondence to: Jan Flusser, E-mail: [flusser@utia.cas.cz](mailto:flusser@utia.cas.cz)

manufacturer is questionable. However, the color-based features remain the main indicators for the classification. The approaches, adopted by various authors, differ from one another only in the color space in which they are working. Traditional color spaces such as RGB and HSV have been commonly used, but some authors proposed specialized color spaces. Most of the methods extract the features at the pixel level,<sup>3–5</sup> others use image blocks,<sup>1</sup> and some methods work with the image as a whole.<sup>6</sup>

The first one who came up with the solution of the blood detection was one of the producers of the capsules, the Israeli medical technology company Given Imaging. Its second-generation capsule had an embedded software for detection of bleeding lesions called the suspected blood indicator (SBI). The SBI detected the “red areas” in the frames and marked the frames as potentially bloody. However, the study by Signorelli et al.<sup>7</sup> shows a low performance of SBI, where the TP, true-negative, FP, and false-negative results, were 40.9%, 70.7%, 69.2%, and 42.6%, respectively. Other studies,<sup>2,8</sup> reached similar conclusions and confirmed that the SBI does not reduce the time required for interpretation of WCE, which was the main goal. This insufficiency initiated designing of new algorithms for detecting blood presence. They can be categorized into three main groups according to the size of the frame area that they are working with.

*Pixel-wise methods* do not work with any local context and process the individual pixels independently. Their main advantages are their simplicity and speed. Some of them work in the RGB color space with spectrum transformation<sup>4</sup> or use a simple thresholding<sup>9</sup> with the threshold found by a support vector machine. However, as we will discuss below in Sec. 3.1, RGB space does not exhibit a good discriminability in tests on more patients. Shahet al.<sup>3</sup> proposed using only the hue component of the hue, saturation, intensity (HSI) representation, and other authors followed their approach. Penna et al.<sup>5</sup> tried to improve the accuracy by edge masking with Mumford–Shah functional; Mohanapriya and Sangeetha<sup>10</sup> used gray-level co-occurrence matrix and classified the pixels by neural networks. In Ref. 6, the color features are extracted with the help of K-means clustering.

*Block-wise methods* divide each frame into blocks of  $n \times n$  pixels and calculate the color features thereof. Each block is evaluated at first separately, then additional criteria for adjacent blocks may be applied.<sup>1</sup> The performance of the block-wise methods is controlled by the block size  $n$ . For large  $n$ , the methods converge to image-wise approach and become robust but lose sensitivity to small blood spots. If  $n$  approaches one, the methods converge to the pixel-wise approach—they become less stable but more sensitive. A choice of  $n = 5$  is recommended as a compromise.

*Image-wise methods* provide global features that describe the whole frame. However, this approach cannot discover small bleeding areas. Still, some authors classified the frames in this way, e.g., Lv et al.<sup>11</sup> used a spatial pyramid of color invariant histograms.

There are of course other image processing methods that were designed for WCE data analysis. For instance, Gueye et al.<sup>12</sup> used WCE for automatic detection of cancer and precancerous colon lesions. They consider polyps, inflammation, tumor, and bleeding areas as suspect regions they wanted to detect. Since the primary features are texture and shape, the authors employed the scale invariant feature transform (SIFT)<sup>13</sup> as the descriptors. Although

they reported promising results in polyp and inflammation detection, their approach cannot be adopted for blood detection in the small intestine. In that case, the texture and the distribution of the key points detected by the SIFT method are not distinctive features for classification. The same is true for the methods that use local binary patterns (LBP) for WCE image description.<sup>14,15</sup> They capture the texture of the intestinal wall, which is not discriminative for bleeding detection.

### 3 Proposed Technique

The aim of this paper is to develop and implement methods for blood detection that recognize blood in the frame regardless of its particular color and size. Special attention is paid to detection of small blood spots (with a diameter around 5 pixels or less), which cannot be detected by most of the earlier methods. We propose two different methods for blood detection, which can be used either individually or their results can be fused by various fusion rules. Let us call these methods A and B in the sequel.

#### 3.1 Method A

Method A works pixel-wise and is based solely on the color. However, it does not work in standard RGB space because it has been known both from the literature<sup>1</sup> as well as from our experiments that the RGB space does not provide enough discriminability of blood pixels (see Fig. 1). We define our new color space such that the separability of blood pixels and intestinal wall should be maximized. The study we performed on 15 patients shows that an appropriate color space can be defined, as shown below in the first step. The complete algorithm can be summarized as follows:

$$1. K = \min(1 - R, 1 - G, 1 - B),$$

$$M = (1 - G - K),$$

where  $R, G$ , and  $B \in \langle 0, 255 \rangle$ . This color space is similar to the popular CMYK space. The pixels with a low value in green and high values in red and blue are well separated.

$$2. R_1 = \sqrt{G^2 + B^2},$$

$$R_n = \begin{cases} 0 & \text{if } R_1 = 0 \wedge R < 128 \\ 255 & \text{if } R_1 = 0 \wedge R \geq 128 \\ R/R_1 & \text{if } R_1 \neq 0 \end{cases}$$

This transform emphasizes the red channel.

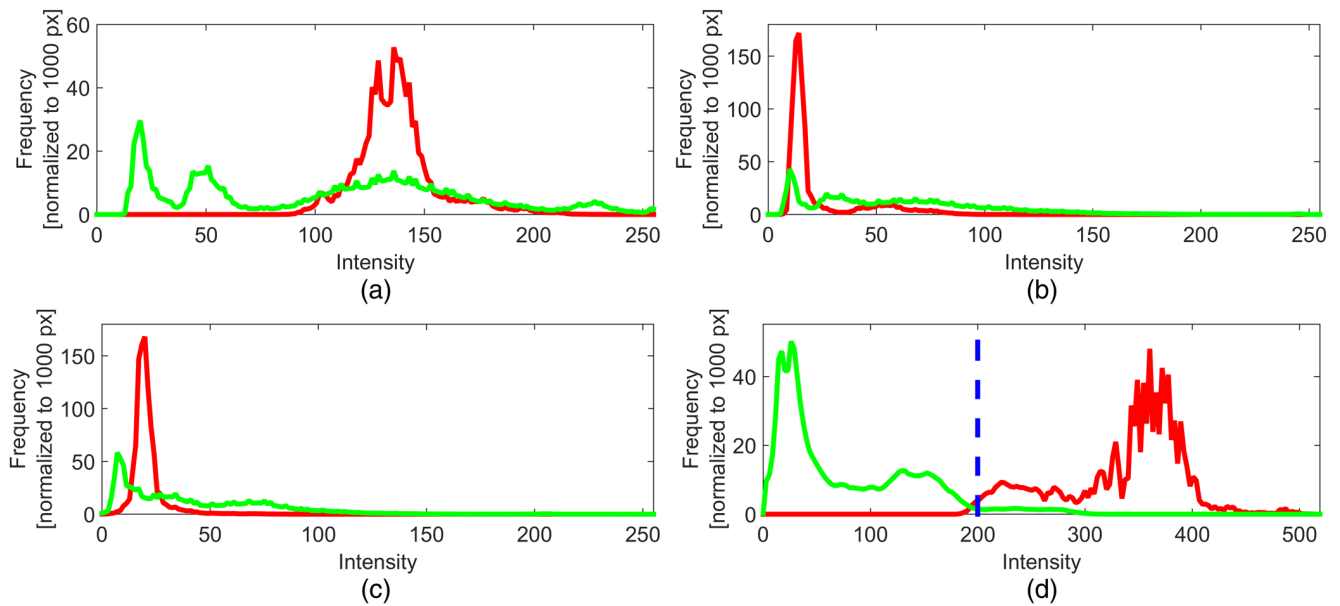
3. The classification criterion  $C$  is defined as

$$C = R_n \cdot M.$$

The number of pixels of the frame, where the  $C$ -value exceeds 200, is denoted as  $N_C$ .

4. Finally,  $N_C$  is compared to a user-defined threshold of “the required number of blood pixels”  $t$ . If  $N_C \geq t$ , the frame is classified as positive (i.e., as one that may contain blood).

The main advantage of method A is its speed because it does not contain any high-level operations. As we will see in the experimental section, the method provides a good TP rate.



**Fig. 1** (a)–(c) The separability of blood (red histogram) against the background (green histogram) in various color bands. The histograms were calculated over several hundred manually selected blood and blood-free patches. None of the RGB channels provides a sufficient discriminability. (d) The new color space separates the blood very well even in one dimension, given by the value of  $C$ . The empirically selected decision threshold on this training data is about 200. This value was used in all experiments in this paper.

### 3.2 Method B

Method B uses a more sophisticated approach that not only is based on pixel colors but also incorporates the assumption that the blood in the frame forms a continuous region (or a few such regions). In other words, it eliminates isolated pixels or small spots with a color similar to blood but that probably do not exhibit actual bleeding. Thanks to this, method B achieves a low FP rate; however, it is at the expense of the computational time. On the other hand, it yields generally higher false-negative rate than A.

The algorithm consists of four main steps:

1. The Canny edge detector<sup>16</sup> is applied to find closed-boundary regions.
2. Morphological erosion is applied to remove small regions. The term “small” is given by a user-defined maximum diameter.
3. The color of the input image is converted to HSV. If the color is within the interval of blood color, defined in advance by training, we mask the respective pixel.

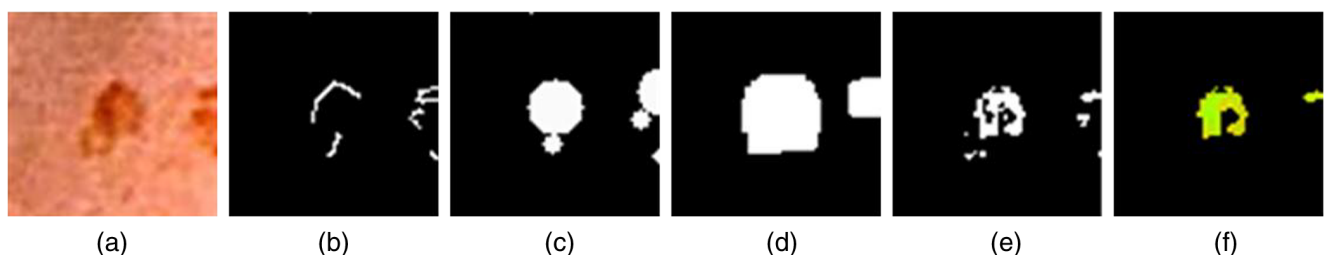
4. The intersection of the outputs of step 2 and step 3 is classified as a blood spot.

For illustration of the individual steps, see Fig. 2.

## 4 Technical Parameters and Implementation

In our study, we were working with two WCE systems—PillCam SB3 by Given Imaging<sup>17</sup> and EndoCapsule developed by Olympus.<sup>18</sup> The former capsule produces a video of a spatial resolution  $256 \times 256$  pixels with variable sample frequency, where the frames per second (fps) rate fluctuates from 0.5 to 3; the latter one yields a  $288 \times 288$  video in the stable rate 2 fps. Both videos are stored in proprietary compression formats based on motion JPEG. We decoded these formats and extracted individual video frames.

Both companies also provide a simple blood detection software along with the capsule. This software apparently works in RGB space, trying to detect the frames containing the “above-threshold” number of pixels of “blood color” (the details about the implemented algorithms are not available). As we already mentioned, the “blood color” varies significantly among the



**Fig. 2** Blood detection by Method B. (a) Input image. (b) Output of Canny detector. (c) Approximate closed-boundary regions. (d) Morphological operation. (e) At the same time, the input image (a) is converted to HSV and potential blood pixels are masked. (f) The output created by intersection of (d) and (e).

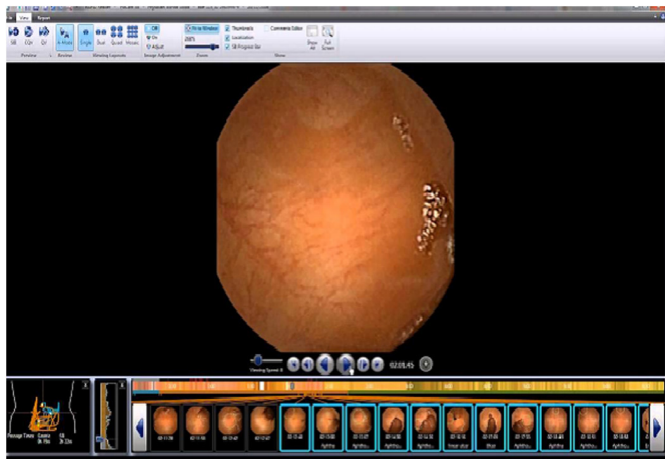
patients, which is the main drawback of these simple algorithms. Nevertheless, both Olympus as well as Given Imaging software are very fast, have comfortable graphical interface [see Figs. 3 (b) and 3(a)], and may be useful for approximative view, but they are not suitable for a rigorous analysis.

Both our methods were implemented in C language and contain a user-friendly graphical user interface (GUI) to provide the doctor with maximum comfort. The input parameter  $t$  can be easily changed by a GUI slider (see Fig. 4 for a GUI screenshot). Note that  $t$  becomes effective in the last step of the algorithm. Hence, the steps 1 to 3 are performed automatically and only

once, regardless of particular value of  $t$ . When changing  $t$ , only a single-number comparison is performed for each frame, and the results are displayed in real time. This is a big advantage of method A.

## 5 Experiments

We tested the performance of the methods on real WCE data. We used videos of 15 patients of the length from 12,000 to 20,000 frames each. The patients suffered from bleeding into the small intestine of various intensities and scales. Some patients bled at more than one place.

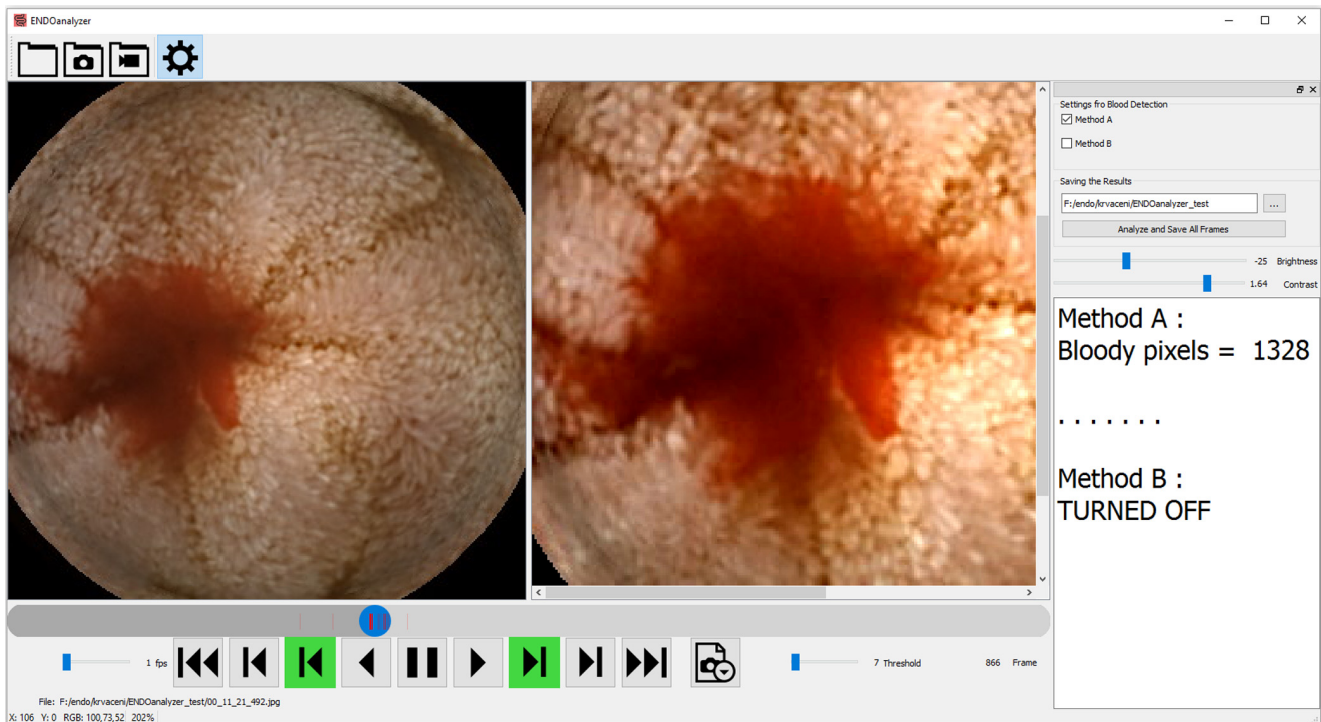


(a)

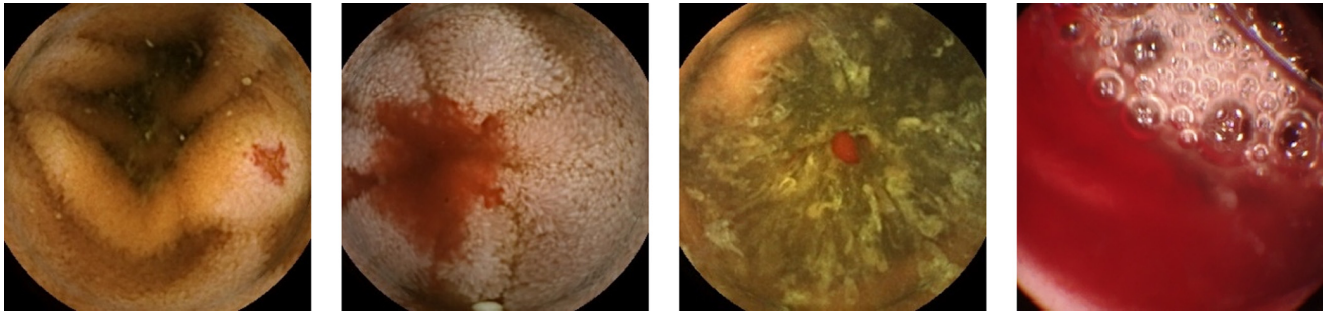


(b)

**Fig. 3** Screen shots of commercial software: (a) Given imaging and (b) Olympus.



**Fig. 4** Screen shot of our software solution. The frames classified as positive are marked by a small red bar in the time axis. The thickness of the bar is proportional to the number of blood pixels. The frames can be visualized for inspection (left) and for interactive contrast enhancement if necessary (right).



**Fig. 5** Examples of frames containing blood annotated by a doctor.

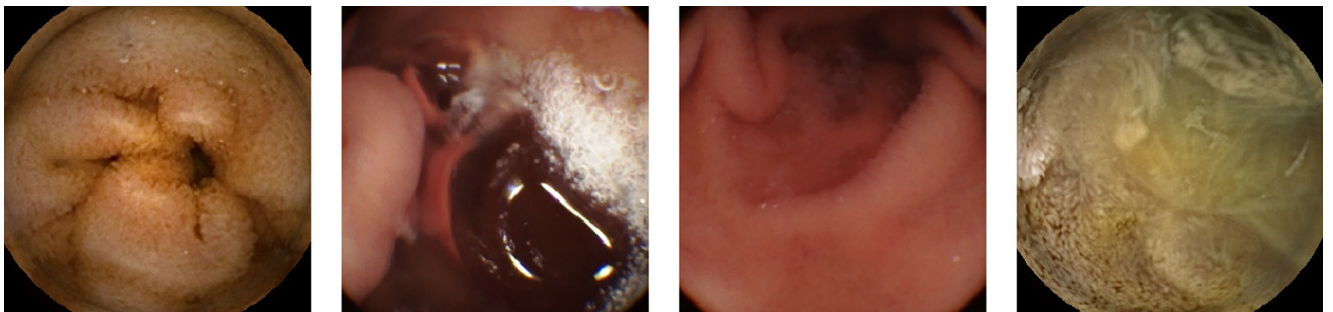
All videos were first annotated manually by an experienced endoscopist, who selected 390 frames altogether of bleeding of various types and extents (see Fig. 5 for some examples). Among them, he marked 339 frames as “serious bleeding.” Then, he selected other 1500 frames with no bleeding (see Fig. 6 for some examples). Some of them are of an appearance that could be misinterpreted as bleeding because of the presence of red spots. We did not use this knowledge for training but solely as a ground truth for evaluating the performance of the methods.

The results are summarized in Tables 1–3. We can see the performance of A and B when used individually along with the performance of various fusion rules. The rule AB means that the frame must be classified as a blood frame by both A and B together.  $A + B$  means that at least one method must classify the frame as positive. The rule  $A - B$  requires the frame to be labeled as positive by A but as negative by B. In the second column of Table 1, we see the TP rate achieved on 339 significant blood frames. The third column shows the same calculated for all 390 blood frames. The last column shows the FP rate

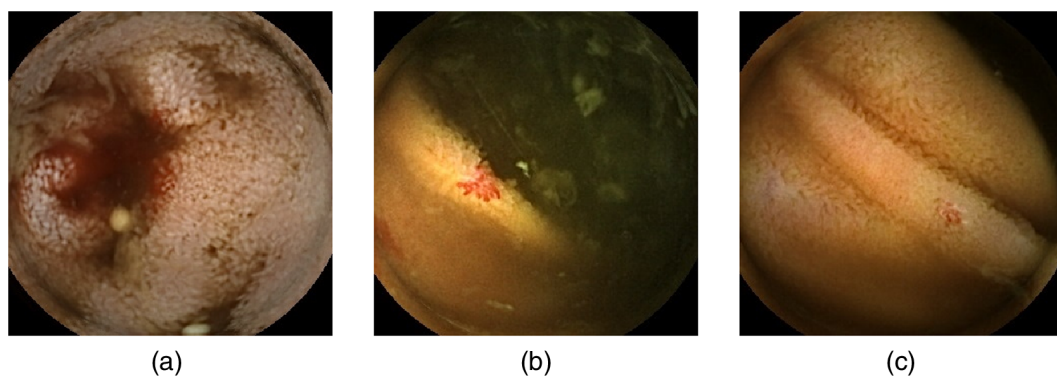
evaluated on the set of 1500 blood-free frames. Both A and B require few user-defined parameters. The main parameter is the number of suspected pixels  $t$  in method A. Table 1 is for  $t = 1$ , Table 2 for  $t = 5$ , and Table 3 for  $t = 10$ .

Based on this experiment, we can deduce the following conclusions.

- Method A, if used individually, yields good TP and FP ratios even if the blood spots are very small (2 to 5 pixels). Method B is individually significantly worse in terms of TPs. If the fusion rule AB has been used, the FP rate falls below 5% but the TP rate does not exceed 60%, which is not a good choice. If the fusion  $A + B$  has been used, then both rates are acceptable by the endoscopists—a TP rate above 80% and a FP rate around 20%. This combination seems to be optimal if the computing complexity (and consequently the time of processing) is not a crucial criterion. If time plays an important role, then individual A is



**Fig. 6** Examples of blood-free frames annotated by a doctor.



**Fig. 7** Examples of hard-to-detect blood frames. (a) Found by A, missed by B. (b) Found by B, missed by A. (c) Not found.

the best choice. On a common PC, it is able to process >80 fps.

- The dependence of the performance of A on the parameter  $t$  is relatively mild. Obviously, both TP and FP rates decrease (with some random fluctuations) as  $t$  increases. Since the TP rate uses time as a primary criterion, we recommend setting  $t$  from 1 to 3 pixels. Note that the method B does not depend on  $t$  at all.
- If the fusion A + B is to be used, the best choice of  $t$  is about 5 pixels.

We performed a comparison with the state-of-the-art method.<sup>19</sup> The TP rate on 390 frames was 58%, whereas the FP rate on 1500 negative frames was 41%. Clearly, our method performs much better in most parameter settings. In Fig. 7 we also show some examples of the frames which are difficult to detect.

Tables 1–3 show our ability to detect bleeding in an isolated frame without taking the time context into account. However, in reality, the appearance of the blood in the intestine is always depicted on more consecutive frames than one. So, we evaluated the results of the same experiments once again with a modified methodology. The detection is marked as “TP” if a ground-truth positive frame lies in its 5-s neighborhood. This corresponds to how the method is actually used—the detected frames are displayed, and the doctor checks their short-time neighborhood. When recalculating the success rates (see Tables 4–6), an excellent performance of our method is clearly demonstrated.

The last statistics illustrates saving of time needed for visual inspection. In the third column of Table 7, we can see the time needed for observation of the original video. The last three

**Table 1** Sensitivity  $t = 1$  pixel.

Method	TP-significant (%)	TP-all (%)	FP (%)
A	86.43	85.64	20.20
A – B	28.91	31.54	15.93
B	58.11	54.62	5.93
B – A	0.59	0.51	1.67
AB	57.52	54.10	4.27
A + B	87.02	86.15	21.87

**Table 2** Sensitivity  $t = 5$  pixels.

Method	TP-significant (%)	TP-all (%)	FP (%)
A	81.12	80.26	15.13
A – B	26.25	28.97	11.27
B	58.11	54.62	5.93
B – A	3.24	3.33	2.07
AB	54.87	51.28	3.87
A + B	84.37	83.59	17.20

**Table 3** Sensitivity  $t = 10$  pixels.

Method	TP-significant (%)	TP-all (%)	FP (%)
A	75.81	74.87	12.80
A – B	22.42	25.13	9.53
B	58.11	54.62	5.93
B – A	4.72	4.87	2.67
AB	53.39	49.74	3.27
A + B	80.53	79.74	15.47

**Table 4** Sensitivity  $t = 1$  pixel—at least one frame within  $\pm 5$  s.

Method	TP-significant (%)	TP-all (%)	FP (%)
A	98.23	96.41	20.20
A – B	11.50	13.59	15.93
B	86.73	82.82	5.93
B – A	0.00	0.00	1.67
AB	86.73	82.82	4.27
A + B	98.23	96.41	21.87

**Table 5** Sensitivity  $t = 5$  pixels—at least one frame within  $\pm 5$  s.

Method	TP-significant (%)	TP-all (%)	FP (%)
A	96.76	94.10	15.13
A – B	10.91	12.31	11.27
B	86.73	82.82	5.93
B – A	0.88	1.03	2.07
AB	85.84	81.79	3.87
A + B	97.64	95.13	17.20

**Table 6** Sensitivity  $t = 10$  pixels—at least one frame within  $\pm 5$  s.

Method	TP-significant (%)	TP-all (%)	FP (%)
A	92.04	89.74	12.80
A – B	8.26	9.74	9.53
B	86.73	82.82	5.93
B – A	2.95	2.82	2.67
AB	83.78	80.00	3.27
A + B	94.99	92.56	15.47

**Table 7** Reducing the inspection time by the proposed method.

Patient	No. of frames	Video time	$t = 1$ (%)	$t = 5$ (%)	$t = 10$ (%)
Pat 2	10,174	56:31	3.47	<b>1.37</b>	0.66
Pat 8	32,750	3:01:56	2.41	<b>1.76</b>	1.51
Pat 9	31,389	2:54:23	4.94	<b>2.14</b>	1.15
Pat 7	11,140	1:01:53	2.84	<b>2.20</b>	1.65
Pat 14	63,629	5:53:29	2.58	<b>2.41</b>	2.32
Pat 12	13,322	1:14:00	6.52	<b>3.44</b>	2.13
Pat 5	19,078	1:45:59	6.81	<b>4.25</b>	3.03
Pat 13	63,992	5:55:30	4.64	<b>4.38</b>	4.22
Pat 4	16,909	1:33:56	9.41	<b>5.81</b>	3.88
Pat 6	25,504	2:21:41	11.72	<b>9.08</b>	7.09
Pat 11	19,742	1:49:40	10.81	<b>9.80</b>	9.17
Pat 1	13,432	1:14:37	14.89	<b>10.36</b>	7.86
Pat 10	29,832	2:45:44	22.83	<b>18.92</b>	16.59
Pat 15	24,762	2:17:34	30.84	<b>27.21</b>	24.93
Pat 3	87,075	8:03:45	90.17	<b>87.28</b>	85.50
Median	24,762	2:17:34	6.81	<b>4.38</b>	3.88

columns show the time spent by a doctor (in percents of the original time) when checking the positive frames only depending on the choice of parameter  $t$  (we assume that the frames are displayed in 3 fps rate). The fifth column with the bold values corresponds to the recommended choice of parameter  $t = 5$ . The time saving, of course, depends on the extent of bleeding, and in our study varied from 1% to almost 90% in the case of a patient with extensive bleeding throughout the digestive tract.

## 6 Conclusion

In this paper, we proposed two automatic methods for detecting bleeding in WCE video of the small intestine. The first one uses solely the color information, whereas the second one incorporates the assumptions about the blood spot shape and size. The original idea is namely the definition of a new color space that provides good separability of blood pixels and the intestinal wall.

We tested both methods individually as well as in various combinations. We evaluated the results on a large test set, manually annotated by an endoscopist. The conclusion is that method A gives good results in terms of both TP and FP rates. The fusion rule A + B further enhances this result because method B is based on different assumptions, but A + B is significantly slower than the individual A.

We compared the proposed method with the method published in Ref. 19, which proved our method superior. Objective comparison with other methods published in the literature is unfortunately not possible because each author uses his own dataset, and different hardware and software platforms; additionally, most authors have not made their codes publicly available.

We believe the presented technique significantly saves the time of endoscopists, which are required for visual expert assessment. The method is currently in use at the University Hospital at Hradec Kralove of the Charles University, Czech Republic.

## Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

## Acknowledgments

The study was supported by the IGA NT 13532-4/2012 research grant provided by the Czech Ministry of Health Care.

## References

1. Y.-G. Lee and G. Yoon, "Real-time image analysis of capsule endoscopy for bleeding discrimination in embedded system platform," *Int. J. Med. Health Biomed. Bioeng. Pharm. Eng.* **5**(11), 583–587 (2011).
2. S. C. Park et al., "Sensitivity of the suspected blood indicator: an experimental study," *World J. Gastroenterol.* **18**, 4169–4174 (2012).
3. S. K. Shah, J. K. Lee, and M. E. Celebi, "Classification of bleeding images in wireless capsule endoscopy using HSI color domain and region segmentation," in *Proc. of 2007 New England American Society for Engineering Education Conf.* (2007).
4. Y. S. Jung et al., "Active blood detection in a high resolution capsule endoscopy using color spectrum transformation," in *2008 Int. Conf. on BioMedical Engineering and Informatics*, Vol. **1**, pp. 859–862 (2008).
5. B. Penna et al., "A technique for blood detection in wireless capsule endoscopy images," in *17th European Signal Processing Conf.*, pp. 1864–1868 (2009).
6. Y. Yuan, B. Li, and M. Q. H. Meng, "Bleeding frame and region detection in the wireless capsule endoscopy video," *IEEE J. Biomed. Health Inform.* **20**, 624–630 (2016).
7. C. Signorelli et al., "Sensitivity and specificity of the suspected blood identification system in video capsule enteroscopy," *Endoscopy* **37**, 1170–1173 (2005).
8. P.-N. D'Halluin et al., "Does the suspected blood indicator improve the detection of bleeding lesions by capsule endoscopy?," *Gastrointestinal Endoscopy* **61**(2), 243–249 (2005).
9. T. Ghosh, S. A. Fattah, and K. A. Wahid, "Automatic bleeding detection in wireless capsule endoscopy based on rgb pixel intensity ratio," in *2014 Int. Conf. on Electrical Engineering and Information Communication Technology (ICEEICT)*, pp. 1–4 (2014).
10. P. Mohanapriya and M. Sangeetha, "An efficient approach to detect bleeding region in gi tract using segmentation and classification techniques," *Int. J. Adv. Inf. Commun. Technol.* **1**(1), 153–159 (2014).
11. G. Lv, G. Yan, and Z. Wang, "Bleeding detection in wireless capsule endoscopy images based on color invariants and spatial pyramids using support vector machines," in *2011 Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 6643–6646 (2011).
12. L. Gueye et al., "Automatic detection of colonoscopic anomalies using capsule endoscopy," in *2015 IEEE Int. Conf. on Image Processing (ICIP)*, pp. 1061–1064 (2015).
13. D. G. Lowe, "Object recognition from local scale-invariant features," in *The Proc. of the Seventh IEEE Int. Conf. on Computer Vision*, Vol. **2**, 1150–1157 (1999).
14. B. Li and M. Q. H. Meng, "Computer-aided detection of bleeding regions for capsule endoscopy images," *IEEE Trans. Biomed. Eng.* **56**, 1032–1039 (2009).
15. Q. Zhao and M. Q. H. Meng, "Polyp detection in wireless capsule endoscopy images using novel color texture features," in *2011 9th World Congress on Intelligent Control and Automation (WCICA)*, pp. 948–952 (2011).
16. J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-8**, 679–698 (1986).
17. Given Imaging, "Rapid for pillcam software," <http://www.givenimaging.com/> (2 May 2016).
18. Olympus, "Endocapsule 10 software client," <http://www.olympus-europa.com/medical/> (2 May 2016).
19. A. A. Al-Rahayfeh and A. A. Abuzneid, "Detection of bleeding in wireless capsule endoscopy images using range ratio color," ArXiv e-prints (2010).

**Adam Novozámský** received his MSc degree in informatics from the Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University, Prague, in 2010. He is currently pursuing his PhD with the Institute of Information Theory and Automation Cooperating Institute of Czech Technical University, Prague. His research interests include medical imaging, image segmentation, and image forensics.

**Jan Flusser** received his MSc degree in mathematical engineering from Czech Technical University, Prague, Czech Republic, in 1985, his PhD in computer science in 1990, his DrSc. degree in 2001, and became a full professor in 2004. Since 1985, he has been with the Institute of Information Theory and Automation, Czech Academy of Sciences (director of the Institute in 2007 to 2017). His research interests include moments and moment invariants, image registration, image fusion, multichannel blind deconvolution, and super-resolution imaging.

**Ilja Tachecí** received his MD degree in general medicine from the Charles University (Faculty of Medicine in Hradec Kralove) in 1999 and his PhD in internal medicine in 2010. He is currently deputy head for Medical Education at 2nd Department of Internal Medicine, University Hospital in Hradec Kralove. His research interest covers invasive and experimental gastrointestinal endoscopy focused on the small bowel diseases.

**Lukáš Sulík** received his MSc degree in applied informatics from the Faculty of Informatics and Management, University of Hradec Kralove, Czech Republic, in 2016. He is currently in postgraduate study at the Center of Basic and Applied Research of the same university. His interests are in image processing.

**Jan Bureš** graduated from Charles University in 1979. He has been affiliated with the 2nd Department of Medicine, Charles University Faculty of Medicine and University Hospital in Hradec Kralove since 1979. He became an associate professor in 1995 and a professor of internal medicine in 2002. He was elected a fellow of the Czech Medical Academy (FCMA) in 2014. Currently, he is the head of the Academic Department of Internal Medicine. His clinical and experimental research is directed at digestive endoscopy, gastrointestinal microbiome, and inflammatory bowel disease.

**Ondřej Krejcar** received his MSc degree in control and information systems from the Technical University of Ostrava, Ostrava, Czech Republic, in 2002, his PhD in technical cybernetics in 2008, and became an associate professor in 2011 in technical cybernetics at the same university. He is currently a vice-dean for science and research at the Faculty of Informatics and Management of the University of Hradec Kralove. His research interests include biomedicine, image segmentation and recognition, video processing, biometrics, technical cybernetics, and ubiquitous computing.

# IMAGE ANALYSIS OF VIDEOKYMOGRAPHIC DATA

Adam Novozámský<sup>a</sup>, Jiří Sedlář<sup>a</sup>, Aleš Zita<sup>a</sup>, Filip Šroubek<sup>a</sup>, Jan Flusser<sup>a</sup>,  
Jan G. Švec<sup>b</sup>, Jitka Vydrová<sup>c</sup>, Barbara Zitová<sup>a</sup> \*

<sup>a</sup>Institute of Information Theory and Automation  
Academy of Sciences of the Czech Republic, Prague, Czech Republic  
{novozamsky,sedlar,zita,zitova,sroubekf, flusser}@utia.cas.cz

<sup>b</sup>Voice Research Lab, Department of Biophysics  
Faculty of Sciences, Palacký University, Olomouc, Czech Republic  
svecjang@gmail.com  
Voice Centre Prague, Medical Healthcom, Ltd., Czech Republic  
vydrova@medico.cz

## ABSTRACT

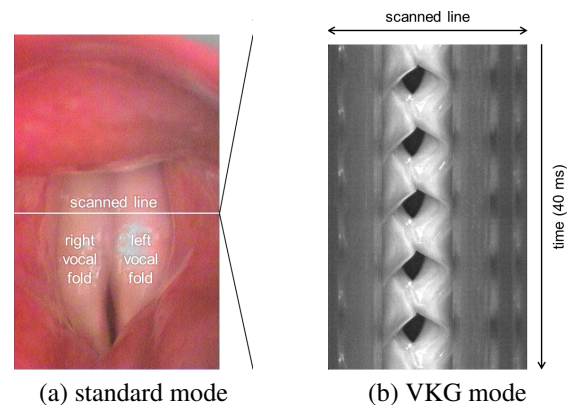
Videokymography (VKG) is a high-speed medical imaging technique used in laryngology and phoniatrics for examination of vocal fold vibrations, it offers important characteristics for diagnosis and treatment of voice disorders. VKG repeatedly scans only a single line from the scene and captures movements of vocal folds in this region of interest. This paper proposes methods for computer assisted evaluation of diagnostically important vibration features, related to movements of vocal folds and their surroundings. They are derived from existing as well as newly developed methods of digital image processing, mainly based on data segmentation and morphological operations. Performance of the developed methods is compared to expert manual assessments and it proves to be comparable with clinicians conclusions.

**Index Terms**— videokymography, medical imaging, data segmentation

## 1. INTRODUCTION

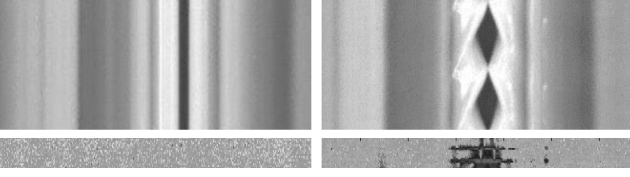
Digital image processing methods form an integral part of medical data analysis and evaluation. Our paper addresses an analysis of videokymographic data (videokymograms), which are collected using videokymography (VKG) - a high-speed imaging technique convenient for observation of vocal fold vibrations. VKG is used in laryngology and phoniatrics for diagnosis of vibration parameters of vocal folds. Our aim is to complement visual evaluation of videokymograms, which can be tedious and clinician-dependent. We proposed automatic software tools for VKG preprocessing and detection of important features.

\*The work was supported by the Technology Agency of the Czech Republic under the project no TA04010877 and by GACR agency project GA13-29225S.



**Fig. 1.** Two modes of videokymographic camera data acquisition: (a) standard and (b) videokymographic. The videokymogram (b) is composed of successively acquired scanned lines at the location indicated in (a).

There are several techniques of capturing human vocal fold vibrations for assessment of their functionality. The most commonly used are the videostroboscopy, high-speed videoendoscopy and the latest, videokymography (VKG). The VKG is an original Czech-Dutch method, developed in 1994 in Groningen (NL) as an alternative to high-speed video recording [1]. The system consists of specially adapted CCD camera, which operates in two modes - standard (50 fps - interlaced, Figure 1(a)) and in high speed (currently 7200 lines per second), when the system records images of a single horizontal line of the selected camera row and stacks them below each other (Figure 1 (b)). The method allows efficient recording of vibration patterns of vocal folds. An application of digital image processing methods for analysis of vocal fold



**Fig. 2.** The VKG data without (top left) and with (top right) an activity. Respective representations in the column-wise Fourier spectral domain close-ups are shown in the bottom line.

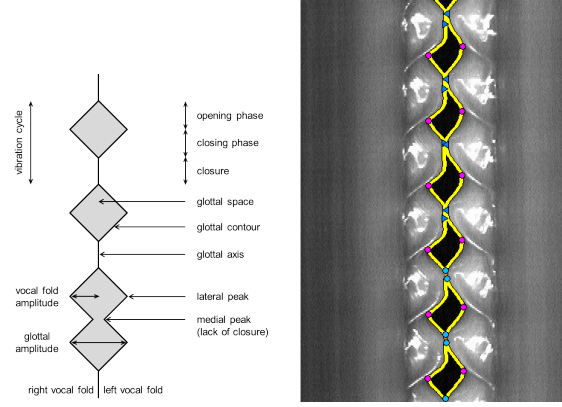
data have been attracting an attention for some time. Most of them have been oriented on high speed videoendoscopy [2], while VKG attracted lesser interest [3], even though this modality offers many benefits due to its efficient data and vocal fold characteristics representation [4]. Our approach broadens proposed methodology for VKG data and makes them more robust to VKG data variability. To facilitate the evaluation of VKG data we focus ourselves on three phases of VKG analysis: (I) data preprocessing, (II) vocal folds characteristics extraction and (III) auxiliary features extraction.

## 2. VKG PREPROCESSING

Typical videokymogram is a gray-scale image capturing several openings and closings of vocal folds (see Figure 1(b)). The data can be noisy, with low contrast and with reflections caused by present mucus. All these factors can negatively influence the performance of further software analysis tools. Moreover, due to the manual examination procedure of the data acquisition, the laryngoscope can be randomly shifted from the optimal position. An important factor which influences the data examination is the patient's discontinued phonation, when the vibrations are missing in VKG data at all.

To ensure the best possible outcome of the automatic analysis of VKG data we apply data preprocessing steps such as median denoising, locally-adaptive contrast enhancement and, if needed, mucus reflection removal using adaptive thresholding followed by diffusion inpainting. The effect of unexpected patient movements and his discontinued phonation is handled by selection of meaningful data subsequences only. This method was developed to select only these parts of VKG recording where the vocal folds are approximately in the center of the VKG image and are active. The Fourier transform of fixed width columns is analyzed and these VKGs with the spectral response under the given threshold are omitted from further processing (see Figure 2).

The attention is paid also to VKG data taken in the standard mode. They can be blurred due to the wrong camera focus and movements of the patient. We proposed to apply multichannel blind deconvolution method [6] to improve the sharpness of the data, even that they are only for visual in-



**Fig. 3.** (left) - Vibration features in videokymograms. (right) - The shape of glottal space and detected base glottal features – opening and closing points (light blue), medial peaks (dark blue), and lateral peaks (magenta).

spection and are not used in the further automatic analysis.

feature	notation
opening points	$O_i$
closing points	$C_i$
lateral peaks	$A_i^R, A_i^L$
medial peaks	$M_i^R, M_i^L$

**Table 2.** Base glottal features in videokymograms (see Figure 3); upper indices  $R$  and  $L$  denote the right and left vocal folds, respectively, and lower index  $i$  denotes the number of the vibration cycle in the videokymogram.

## 3. VOCAL FOLD CHARACTERISTICS

Proposed methods for analysis of VKG data are based on vocal folds / glottal contours and detected base features (see Figure 3) which are key elements for computation of established vibration parameters [5]. The primary step for all further VKG evaluation is the detection of glottal contour (Figure 3 - (right), yellow curves), which is realized by means of an thresholding segmentation with an optimized threshold estimated by normalized graph cuts [7, 8]. This approach maximizes dissimilarity between two parts of the scene according to both spatial and gray-level relations of their pixels. The detected glottal space is then used for estimation of elementary glottal features - opening and closing points, lateral and medial peaks, vibration cycles and their opening, closing, open and closed phases, and glottal and vocal fold amplitudes (see Figure 3). The opening and closing points, and the lateral and medial peaks (see Table 2) are the base features and are used for derivation of the other aforementioned features called derived glottal features. Their derivation from the base features is listed in Table 1. They all are used for computation of es-

feature	notation and definition
generalized opening points	$\tilde{O}_i^j = \{O_i, M_i^j\}$
generalized closing points	$\tilde{C}_i^j = \{C_i, M_i^j\}$
opening phase duration	$t_i^{oj} = A_i^j(y) - \tilde{O}_i^j(y)$
closing phase duration	$t_i^{cj} = \tilde{C}_i^j(y) - A_i^j(y)$
open phase duration	$T_i^{oj} = t_i^{oj} + t_i^{cj} = \tilde{C}_i^j(y) - \tilde{O}_i^j(y)$
closed phase duration	$T_i^{cj} = \tilde{O}_{i+1}^j(y) - \tilde{C}_i^j(y)$
vibration cycle duration	$T_i^j = T_i^{oj} + T_i^{cj} = t_i^{oj} + t_i^{cj} + T_i^{cj} = \tilde{O}_{i+1}^j(y) - \tilde{O}_i^j(y)$
vocal fold amplitudes	$a_i^j = \max( A_i^j(x) - \tilde{O}_i^j(x) ,  A_i^j(x) - \tilde{C}_i^j(x) )$
glottal amplitudes	$a_i = A_i^L(x) - A_i^R(x)$

**Table 1.** Derived glottal features in videokymograms [5]; upper index  $j \in \{R, L\}$  denotes the right and left vocal folds, respectively, and lower index  $i$  denotes the number of the vibration cycle in the videokymogram.

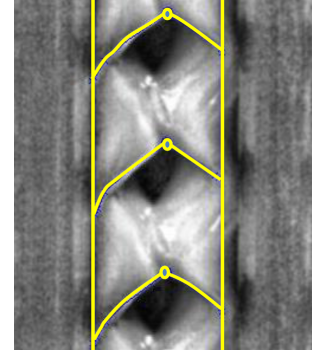
tablished vocal fold vibration parameters [5]. Their detailed definition and discussion can be found in [9].

The proposed base and derived glottal features and set of vocal fold vibration parameters [9, 5] were evaluated on the testing dataset of 50 videokymograms and compared to manual evaluations [10], done by evaluators with different level of experience and resulting in 18 assessment sets in total ( $18 \times 50$  videokymograms). The comparison was realized in an automatic-visual and visual-visual manner. The method introduces the following notation. Let  $P$  denotes the set of both automatically and visually evaluated parameters,  $n$  the number of evaluated videokymograms, and  $m$  the number of visual evaluations. Let  $E_A(p; i) (p \in P; i = 1; \dots; n)$  denotes the result of automatic evaluation of parameter  $p$  in videokymogram  $i$ , and  $E_V(p; i; j) (p \in P; i = 1; \dots; n; j = 1; \dots; m)$  the result of  $j^{th}$  visual evaluation of parameter  $p$  in videokymogram  $i$ . Let  $V^+(p; i)$  denote set of indices of visual evaluations of parameter  $p$  in videokymogram  $i$  with defined result (non-NA)

$$V^+(p; i) = \{j \mid j \in \{1 \dots m\} \wedge E_V(p; i; j) > 0\}$$

Then the consensus result is defined as NA if and only if the result of at least half of corresponding visual evaluations was NA; otherwise, the definition estimates it by the most frequent non-NA result. For each parameter  $p \in P$  and videokymogram  $i \in \{1, \dots, n\}$  the proposed method compares the consensus result of visual evaluations  $E_V(p; i)$  with the result of automatic evaluation  $E_A(p; i)$  (automatic-visual match) and with the results of visual evaluations  $E_V(p; i; j), (j = 1, \dots, n)$  (visual-visual match).

The automatic-visual match compared the automatically estimated parameter categories with the category most frequently selected by the visual evaluators whereas the visual-visual match estimated reliability of the visual evaluations (how often the assessments of the visual evaluators were in agreement). In all cases two evaluations are set to be matching if their respective results fall into the same category or into directly neighboring non-NA categories. The results can be seen in Table 3. Figure 5 illustrates the variability of the



**Fig. 4.** Detected lateral mucosal waves with the starting points (circles) and their detected extent.

VKG data that the proposed algorithms must be able to cope.

The experiments showed consistency between automatic and visual evaluations. The similarity in comparative statistics demonstrates that the performance of the automatic evaluation is comparable with visual evaluations and thus the proposed approach in the computer-aided evaluation is found applicable in clinical practice.

#### 4. LATERAL MUCOSAL WAVES EXTRACTION

Besides of the established set of vibratory features [5] the research was focused on the auxiliary characteristics of vocal chords and their vibrations – laterally travelling mucosal waves. Mucosal waves are tissue waves propagating across located on the upper surface of vocal folds. They propagate laterally across the surface until they disappear or reaches the lateral border of the vocal fold. Their presence and extent can indicate how pliable a vocal fold is and can may indicate problems with stiffness of vibrating tissue.

In videokymograms, mucosal waves are demonstrated as diagonal, sometimes slightly bended lines on vocal folds running in the direction of the opening movement. Their detec-

vibration parameter	automatic–visual match	visual–visual match
NumberOfCyclesR	98%	95%
NumberOfCyclesL	98%	96%
VariabilityR	92%	93%
VariabilityL	88%	93%
ClosureDuration	98%	93%
AmplitudeDifferences	100%	88%
FrequencyDifferences	100%	95%
PhaseDifferences	88%	85%
AxisShift	88%	79%
SkewingR	86%	87%
SkewingL	90%	85%

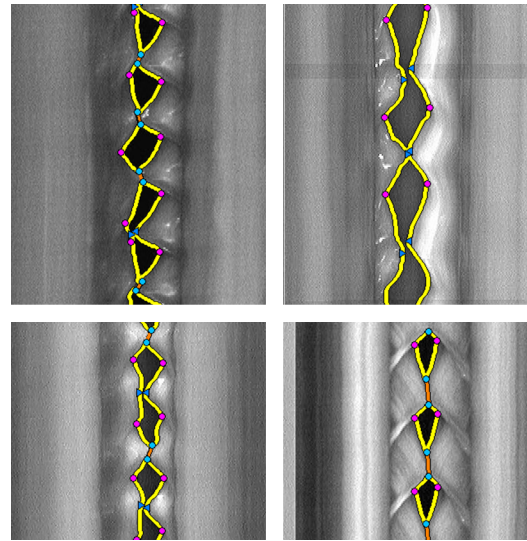
**Table 3.** Comparison of results of automatic and visual evaluations on a set of 50 videokymograms by the automatic–visual and visual–visual match with tolerance between closely neighboring classes. The similarity in comparative statistics for each parameter indicates that the performance of the automatic evaluation is comparable and often better than visual evaluations.

tion can be complicated by reflections, low contrast and their small extent. To solve the problem we introduced the *iterated masked cross-correlation* method. It is based on the detection of self-similarity of the data around the expected position of the wave, which starts at lateral peaks and runs in the direction of the connector of the opening points and the lateral peaks.

The respective cross-correlation kernel is established, positioned at the beginning of the wave and shaped as a tilted rectangle with its size proportional to the glottal space. The kernel is then iteratively updated as the cross-correlation is processed in the given direction, till any progress is made or the steps are shorten below certain level. In this way an approximate path of a glottal wave is constructed and its main direction is then estimated using Fourier spectral analysis of the detected wave path. In Figure 4 there is an illustration of the VKG data with the detected directions and extent of the lateral mucosal waves (circles are representing starting points of the method).

## 5. CONCLUSION

New image processing methods for analysis of vocal fold vibrations in videokymograms were developed. The motivation was to create the tools for the computer-aided evaluation of vibratory patterns to be used by laryngologists and phoniatricians in clinical practice for more detailed diagnosis of voice disorders. The introduced algorithms provide data pre-processing, detection of the glottal space by normalized graph cuts thresholding and extraction of glottal vibration features and parameters proposed in [5]. Comparison of the performance of the developed methods with subjective visual evaluations indicated good match making them promising for future use in clinical practice. In addition to the glottal features an original methodology was also developed for detecting laterally travelling mucosal waves. These features which are revealing on the pliability of the vocal folds and their sur-



**Fig. 5.** The variability of the VKG data with detected features.

rounding are detected using *iterated masked cross-correlation* method. In the future the research will be focused on new sets of auxiliary features describing the close surrounding of vocal folds and their interpretability with respect to the diagnosis and treatment of voice disorders.

## 6. REFERENCES

- [1] J. G. Svec and H. K. Schutte, “Videokymography: high-speed line scanning of vocal fold vibration,” *J Voice*, vol. 10, no. 2, pp. 201–205, 1996.
- [2] J. Lohscheller, U. Eysholdt, H. Toy, and M. Dollinger, “Phonovibrography: Mapping high-speed movies of vocal fold vibrations into 2-d diagrams for visualizing and analyzing the underlying laryngeal dynamics,” *Medi-*

- cal Imaging, IEEE Transactions on*, vol. 27, no. 3, pp. 300–309, March 2008.
- [3] Q. Qiu, H. K. Schutte, Gu L., and Q. Yu, “An automatic method to quantify the vibration properties of human vocal folds via videokymography,” *Folia Phoniatr Logop*, vol. 55, no. 3, pp. 128–136, 2003.
  - [4] J. G. Svec and H. K. Schutte, “Kymographic imaging of laryngeal vibrations. current opinion,” *Otolaryngology & Head and Neck Surgery*, vol. 20, no. 6, pp. 458–465, 2012.
  - [5] J. G. Svec, F. Sram, and H. K. Schutte, “Videokymography in voice disorders: what to look for?,” *The Annals of otology, rhinology, and laryngology*, vol. 116, no. 3, pp. 172180, March 2007.
  - [6] F. Sroubek and J. Flusser, “Multichannel blind deconvolution of spatially misaligned images,” *Image Processing, IEEE Transactions on*, vol. 14, no. 7, pp. 874–883, July 2005.
  - [7] W. Tao, H. Jin, Y. Zhang, L. Liu, and D. Wang, “Image thresholding using graph cuts,” *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 38, no. 5, pp. 1181–1195, Sept 2008.
  - [8] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
  - [9] J. Sedlar, “Image analysis in microscopy and videokymography,” *Ph.D. thesis, Charles University, Prague, Czech Republic*, 2012.
  - [10] V. Hampala, “Visual evaluation of videokymographic features in voice disorders (in czech),” *Master’s thesis, Palacky University, Olomouc, Czech Republic*, 2011.