Video superresolution in real forensic cases

Jan Flusser*, Filip Šroubek†, Jan Kamenický‡ and Barbara Zitová§
Czech Academy of Sciences
Institute of Information Theory and Automation
Pod Vodarenskou vezi 4, 180 00
Prague, Czech Republic
Email: *flusser@utia.cas.cz, †sroubekf@utia.cas.cz, ‡kamenik@utia.cas.cz, §zitova@utia.cas.cz
ORCID: *0000-0003-3747-9214, †0000-0001-6835-4911, §0000-0003-0110-3741

***Abstract*—Visual data, such as images and videos, are frequently used as evidence in court trials. If the data quality is insufficient to convince the court, a carefully tailored data processing algorithm supported with expert's opinion is necessary. We present two real cases from our forensic expertise practice, in which we demonstrate a successful application of video superresolution that helped to convict offenders. The most important feature of image processing algorithms to be legally accepted by the court, is to rule out artifacts with realistic details, which are known to appear for example in deep learning methods.**

***Index Terms*—superresolution, image enhancement, video processing, image/video forensics**

I. Introduction

In forensic image and video processing, the two most common challenges are the following ones: - to enhance image/video quality in order to make it more suitable for visual investigation - to authenticate the image/video data, which means to verify the originality or to identify manipulations/forgeries.

The former requirement typically appears in the stage of police investigation, when the goal is the identification of persons and recognition of details of the crime scene. If the image/video was captured by a low-resolution camera and in poor light conditions, image processing techniques may improve the visual quality substantially.

The latter challenge usually arises if the image/video is presented as an evidence in court. The origin of the visual data source is often questioned and the forensic expertise is required before the data are accepted as an evidence.

This paper is devoted to the former situation. We review two recent criminal cases from our practice[1] where a video was the key evidence but had required a careful computer processing before it was accepted by the court. The main tool applied in both cases is video superresolution, which is a technique that increases spatial resolution, enhances visual quality and recovers details not visible before.

In the next section, we briefly present the superresolution algorithm we developed. In Section 3 and 4, we describe two real criminal cases from the Czech Republic where this technique was successfully used to convict offenders.

II. Superresolution

Superresolution (SR) is a mathematical and software technique which allows us to estimate images with the spatial resolution beyond physical parameters (sensor resolution) of the camera (see [1] for a survey and further references). This technique originally appeared in 1970's to enhance satellite and aerial images. The core idea is quite simple. Let us assume a static scene, which was captured several times with a camera with resolution below the Nyquist frequency of the image. Mathematically, the image formation model is given by equation

$$g_i = D_i u + n_i \,, \tag{1}$$

where g_i's are captured images, D_i's are downsampling operators modelling the camera capturing process, u is the original high-resolution scene, and n_i is additive noise. Assuming there is a random sub-pixel shift between the acquisitions (included in D_i's operators) and that we are able to find the shift parameters, we can register the images with a subpixel accuracy and resample the image on the grid with a spatial resolution higher than that of the inputs; see the illustration in Fig. 1.

However in practice, this produces just rough enhancement. Real input images are (in addition to the under-sampling) degraded by blur, which cannot be removed by SR but should be removed by deconvolution algorithms. In [2], we proposed a method that integrates these two steps together and accomplishes them in a iterative minimization of the functional

$$\min_{u,\{h_i\}} \sum_i \|g_i - DH_i u\|^2 + R(u) + Q(\{h_i\}) \tag{2}$$

where $R(\cdot)$ and $Q(\cdot)$ are image and blur regularization terms, respectively. The image formation model was enhanced by splitting the downsampling operator into the static part D, which performs downsampling with a user-defined factor

This work was supported in part by the Ministry of the Interior of the Czech Republic grant VJ02010029 "AISEE", the Czech Science Foundation grant GA21-03921S, and the *Praemium Academiae* awarded by the Czech Academy of Sciences.

[1]The Institute of Information Theory and Automation (UTIA) is certified by the Czech Ministry of Justice to provide forensic expertise in computer science with special focus on image and video processing and analysis.

979-8-3503-3607-8/23/$31.00 ©2023 IEEE

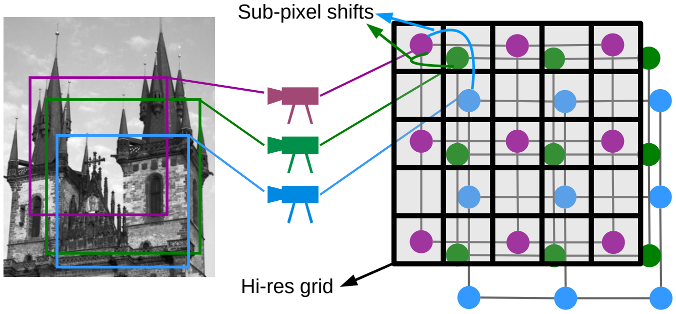

Fig. 1. The basic scheme of still image superresolution: Multiple images capture the same scene slightly shifted. After registering the images with sub-pixel accuracy, we can estimate the high-resolution image by interpolating on a high-resolution grid.

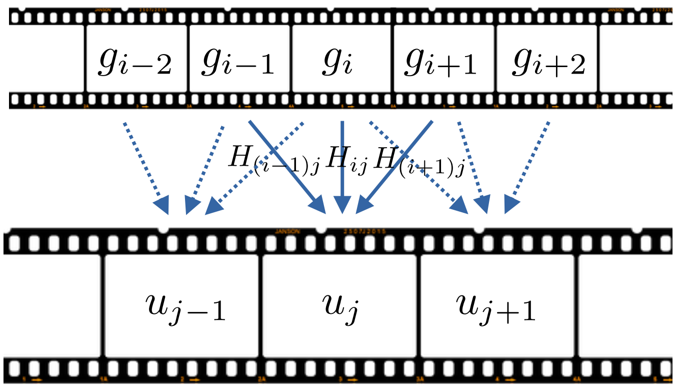

Fig. 2. Video superresolution: A few (in this example three) consecutive low-resolution frames g_i's generate a single superresolved frame u_j. Then the time window is shifted by one frame to generate the next superresolved frame u_{j+1}. H_{ij} denotes the convolution operator that models blur in the i-th frame and shift between the i-th and j-th reference frame.

(e.g. $2\times$ or $3\times$), and the unknown part H_i, which performs convolution with an unknown blur h_i. The sub-pixel shift is modeled by shifting the blur and thus by estimating the blur the shift is seamlessly estimated as well.

This method of simultaneous deconvolution and superresolution proved to be very powerful not only for static images but also for video enhancement. In video processing, we do not need several input channels. We just take few consecutive low-resolution frames to generate a single superresolved frame. Then we shift the "time window" in the input by one frame and generate another superresolved frame (see Fig. 2). Finally, we end up with a superresolved video of the same length and frame rate as the input [3].

Originally, we developed this video-superresolution for thermal imaging cameras, where the resolution is limited by physical constraints (our algorithms is patent-protected in the U.S.A. [4]). However, a similar algorithm can be applied to surveillance/security cameras that operate in the visible band. To save the storage capacity and to keep the price low, these cameras use to have resolution which is too low to capture the details, important in forensic analysis. In practice geometric alignment of video frames is more complex than global shifts and dense motion fields estimated by optical flow are needed. To this end, we developed software "PIZZARO" [5] that implements all the above mentioned methods. As we demonstrate on real cases in the next two sections, the SR method can sometimes recover the details important for the crime investigation.

III. Case 1 – The machete fight

This case happened in August 7, 2011, in the city of Novy Bor and became the most covered criminal case in the Czech Republic in the last decade. A five-man gang attacked guests of the pub/casino after a minor conflict between one of the members of the gang and the pub staff. The attack was swift and extremely brutal. Some members of the gang were armed with machetes, the others "only" with sticks. The fight took about two minutes and resulted in three seriously injured men.

The pub was equipped with four security cameras of very poor quality with partially overlapping fields of view. They had low resolution, high compression, low frame rate (4 fps) and were not time-synchronized. Fortunately, police captured all aggressors so there was no need for face recognition. The main challenge was to identify who of the attackers used a machete and who used a stick. The prosecutor presented a significant difference in a crime classification. The attackers using machetes faced charges of attempted murder and the others only of aggravated assault and battery. As one may expect, all defendants claimed they have no machete. There was no other evidence except the video, which, however, was not accepted by the judge because it did not carry any reliable information; see Fig. 3 for sample frames (original uncut videos of the attack will be presented at the conference). At this moment, when the court trial was hitting a dead end, our Institute was asked to analyze the video and try to answer the question who of the attackers had a machete and who had a stick in his hand.

First, we synchronized all four videos in time such that it was possible to visually track the persons moving from one visual field to another; see Fig. 3. On selected parts of the video we applied superresolution. The superresolved video, along with extracted single frames, was presented at the court (and will be presented at the conference) and was accepted as the key evidence. Several examples of processed frames are in Fig. 4. It was possible to clearly identify guys attacking with machete; see Fig. 5. During the trial, the main objection of the advocacy was that superresolution might introduce artifacts. We had to explain to the court the main principles of the superresolution algorithm in a detailed manner and to convince them that no artifacts could be present. Most probably, if the method used other data for training and for some kind of "superresolution inpainting", the video would be rejected by the court and the truth would remain hidden forever.

Thanks to our video processing, the two attackers with machete were found guilty of the murder trial and were sentenced to 17.5 and 16 years in prison, the others got from 3 to 5 years of imprisonment.

Fig. 3. Time synchronization of four videos capturing the scene from different angles.

IV. Case 2 – The neighborly murder

The second case is the murder in the village of Volfartice (Northern Czechia) that happened in 2020. The case started as a altercation between two neighbors, let us call them Family A and Family B. The man A and the women B started arguing at the gate of the garden surrounding the house of Family B. The conflict rapidly accelerated but was still mainly verbal because both persons were not armed. Then the man B came with a gun to "help" his wife. He shot five times the man A, who died on the spot. The women A recorded the entire story on her smartphone from a window of her house, which means from the distance between 40 and 50 meters.

Although the whole story looked clear for the police, at the court trial the advocate of Family B came up with the construction that the shooting was a legitimate self-defence, because the man A allegedly attacked both persons B (this hypothesis was presented despite the fact that the police did not find any weapon belonging to the man A). Since there were no other witnesses, the video became the key evidence. However, the original video is of such a low quality that it does not clearly show what actually happened; see examples of two frames in Fig. 6(top). First, we performed software stabilization of the video. On the region of interest, we applied contrast and color enhancement and increased the resolution twice. Then we pasted this processed region of interest into the original video to put it into the context; see Fig. 6(bottom) for examples (the complete video will be presented at the conference). The processed video clearly disproves the version about the self-defence of the shooter. It is apparent that the man A had no weapon and did not attack the man B at all. The trial is still running but thanks to the video processing the shooter will be most probably found guilty of murder.

V. Conclusion

In this paper, we presented two real cases from our practice where superresolution enhanced the visual quality of the video presented at the court as an evidence and, consequently, the processed video convicted the offender.

Our 15-years experience in forensic image/video analysis indicates, that careful preprocessing/enhancement for a better visual interpretation is the main issue in most cases. Fully automatic object recognition and scene analysis is seldom required. The methods used for forensic purposes should be deterministic, repeatable, clearly explainable, and well established in the research community. If one uses a new method which is not generally known and does not have a sufficiently long list of successful applications, one must be prepared that the method may be a subject of discussion at the court and of objections of the opponents. If the presented expert's opinion is not convincing enough, then there exists a high risk of rejecting all processed data and conclusions obtained by this method.

In almost all cases, we were questioned whether or not the method may introduced artifacts. The answer "yes" would almost surely lead to rejection, even if the probability of the artifact appearance would be low and the influence on the image/video content would actually be negligible. For similar reasons, the court would most probably reject any data processed using deep learning techniques that rely on training on external data not provided by the expert and not related to the particular case. Although this is an approach widely used in pre-trained CNN's and accepted in the research community, it is not acceptable for forensic purposes (at least not currently in the Czech Republic). The tasks such as noise suppression, superresolution, image deconvolution, inpainting-based restoration, etc. are usually resolved by deep learning very effectively but the dependence on the training set, the risk of introducing artifacts from the training data, and the limited possibility to explain how and why the method generated exactly this result make deep learning not plausible to non-technicians.

References

[1] P. Milanfar, Ed., *Super-resolution Imaging*. CRC Press, 2010.

[2] F. Šroubek, G. Cristóbal, and J. Flusser, "A unified approach to superresolution and multichannel blind deconvolution," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2322–2332, 2007.

[3] F. Šroubek, J. Flusser, and M. Šorel, "Superresolution and blind deconvolution of video," in *Proceedings on the 19th International Conference on Pattern Recognition*. IEEE, 2008, pp. 1–4.

[4] M. Stratmann, J. Evers-Senne, M. Schmieder, J. Flusser, and F. Šroubek, "Method for preparing images in non-visible spectral ranges, and corresponding camera and measuring arrangement," U.S. Patent 9 386 238, 2016.

[5] J. Kamenický, M. Bartoš, J. Flusser, B. Mahdian, J. Kotera, A. Novozámský, S. Saic, F. Šroubek, M. Šorel, A. Zita, B. Zitová, Z. Šíma, P. Švarc, and J. Hořínek, "Pizzaro: Forensic analysis and restoration of image and video data," *Forensic Science International*, vol. 264, no. 1, pp. 153–166, 2016.

Fig. 4. Machete fight: (top) Selected frames of the original video. (bottom) Examples of superresolved frames

Fig. 5. Final superresolved frames. The machete and stick are clearly visible.

Fig. 6. Neighbor dispute shooting: (top) Two frames of the original video. (bottom) Superresolution of regions of interest. (left) A scuffle between the man A and the woman B. (right) Shooting of the man A by the man B.