Noise-Resistant Image Retrieval

Cyril Höschl IV and Jan Flusser,
The Institute of Information Theory and Automation of the ASCR,
Pod vodárenskou věží 4, 182 08 Praha 8, Czech Republic
Email: flusser@utia.cas.cz

***Abstract*—We present a content-based image retrieval method which is particularly designed for noisy images. The images are retrieved according to histogram similarity. To reach high robustness to noise, the histograms are described by novel features which are insensitive to convolution with a Gaussian kernel, i.e. insensitive to a Gaussian additive noise in original images. The advantage of the new method is demonstrated experimentally on real data.**

I. Introduction

Since the appearance of the first image databases in the 80's, image retrieval has been a goal of intensive research. Early methods did not search the image themselves but utilized some kind of metadata and image annotation to retrieve the desired images. As many large-scale databases do not contain any annotation, content-based image retrieval (CBIR) methods have became one of the most important challenges in computer vision. By CBIR we understand methods that search a database and look for images which are the "most similar" (in a pre-defined metric) to a given query image. CBIR methods do not rely on text annotation and/or other metadata but analyze the actual content of the images. Each image is described by a set of features (often hierarchical or highly compressive ones), which may reflect the image content characteristics the user prefers – colors, textures, dominant object shapes, etc. The between-image similarity is then measured by a proper (pseudo)metric in the corresponding feature space.

CBIR is a subjective task because there is no "objective" similarity measure between the images. Hence, many CBIR systems aim to retrieve images which are perceived as the most similar to the query image for majority of users and the users feel this similarity at the first sight without detailed exploration of the image content. This requirement, along with a need for a fast system response, has lead to a frequent utilization of low-level lossy features based on image colors/graylevels. A typical example is an intensity or color histogram. It is well known that histogram similarity is a salient property for human vision. Two images with similar histograms are mostly perceived as similar even if their actual content may be very different and, on the other hand, the images having substantially different histograms are rarely ranked by the observers as similar. Another attractive property of the histogram is that it does not depend on image translation, rotation and scaling (normalized to image size), and depends only slightly on elastic deformations. Thanks to this, one need not care about image geometry and look for geometric invariants. Simple preprocessing can also make the histogram insensitive to linear variations to contrast and brightness of the image. Hence, histogram established itself as a meaningful image characteristic for CBIR [7], [9], [8].

Histogram is rarely used for CBIR directly as it is basically for two reasons. The histogram is not only an inefficiently large structure (in case of color images, the RGB histogram is stored in a vector of 2^{24} integers, which may be even more than the memory requirement of the original image) but also uselessly detailed. It is sufficient and computationally efficient to capture only prominent features of the histogram and suppress insignificant details. To do so, some authors compressed the histogram from the full color range into few bins [3], [4] while some others represent the histogram by its coefficients in a proper functional basis. The advantage of the latter approach is that the number of coefficients is a user-defined parameter – we may control the trade-off between high compression on one hand and accurate representation on the other hand. It is very natural to get inspired by a clear analogy between histogram of an image and probability density function (pdf) of a random variable. In probability theory, the pdf uses to be characterized by its moments, so it is worth applying the same approach in histogram-based CBIR [6], [10].

CBIR methods based on comparing histograms are sensitive to noise in the images, regardless of the particular histogram representation. Additive noise results in a histogram smoothing, the degree of which is proportional to the amount of noise. This immediately leads to drop of the retrieval performance because different histograms tend to be more and more similar to each other due to their smoothing. In digital photography, the noise is unavoidable. When taking a picture in low light, we use high ISO and/or long exposure. Both amplifies the background noise, which is present in any electronic system, such that the noise energy may be even higher than that of the signal. Particularly compact cameras and cell-phone cameras with small-size chips (i.e. devices which produce vast majority of photographs on Flickr, on other servers and on personal websites) suffer from this kind of noise, along with an omnipresent thermal noise. In-built noise reduction algorithms are able to suppress the noise only slightly and perform on the expense of fine image details.

Although the noise in digital photographs is an issue we cannot neither avoid nor ignore, very little attention has been paid to developing noise-resistant CBIR methods. The authors of the papers on CBIR either skip this problem at all or rely on denoising algorithms applied to all images before they enter the database. Such solution is however not convenient or even not realistic, because the denoising inevitably introduces artifacts such as high-frequency cut-off, requires additional time and mostly also needs a cooperation of the user in choosing proper

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.513

parameters. In this paper, we present an original histogram-based image retrieval method which is not only robust but totally resistant (at least theoretically) to additive noise. The core idea of the method is a proper representation of the histogram by certain characteristics, which are not affected by image noise. We stress that the paper is *not* aimed to judge in which tasks and for what purposes a histogram-based CBIR is appropriate. We rather show if it is appropriate, how it should be implemented in case of noisy database and/or query images.

II. The noise model

As we already mentioned, we primarily consider the thermal noise and electronic noise of consumer cameras. It is a common belief that such noise n can be modelled as stationary additive Gaussian white noise (AGWN) with zero mean and standard deviation σ, and that the noise is not correlated with the original image f.[1] We adopt this assumption in this paper, too. Hence, the noise histogram h_n has a Gaussian form

$$h_n(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{t^2}{2\sigma^2}\right), \tag{1}$$

where t is the index of the graylevel. The histogram h_g of the noisy image $g = f + n$ is then a convolution of the original histogram and the noise histogram

$$h_g(t) = (h_f * h_n)(t).$$

All histograms in this paper are normalized w.r.t. the image size such that they have a unit integral.

III. Histogram representation resistant to noise

In this Section, we present a representation of image histogram by descriptors which are not affected by AGWN. These descriptors are based on the statistical moments of the histogram, which is in probability theory a common approach to characterization of pdf's. Let h be a pdf of a random variable X. Then the quantity

$$m_p^{(h)} = \int x^p h(x) dx \tag{2}$$

where $p = 0, 1, 2, \cdots$, is called *p-th order general moment* of X. Clearly $m_0 = 1$ and m_1 equals the mean value of X. In general, the existence (i.e. finitness) of the moments is not guaranteed, however if h is a (normalized) histogram, its support is bounded and all m_p's exist and are finite. On the other hand, any compactly-supported pdf can be exactly reconstructed from the set of all its moments.In this sense the moments provide a complete and non-redundant description of a pdf/histogram.

Unfortunately, the moments of the histogram themselves are affected by image noise. As the histogram of the noisy image is a smoothed version of the original histogram, it holds for its moments

$$m_p^{(g)} = \sum_{k=0}^{p} \binom{p}{k} m_k^{(n)} m_{p-k}^{(f)}. \tag{3}$$

[1]There exist also other components of image noise which are not Gaussian, namely Poisson-type photon shot noise and quantization noise. Shot noise is usually supposed to be close to Gaussian while quantization noise is disregarded.

This assertion can be easily proven just using the definitions of moments and of convolution. Since the noise is supposed to be zero-mean Gaussian, h_n has a form of (1) and its moments are

$$m_p^{(n)} = \sigma^p (p-1)!! \tag{4}$$

for any even p. The symbol $k!!$ means a double factorial, $k!! = 1 \cdot 3 \cdot 5 \cdots k$ for odd k, and by definition $(-1)!! = 0!! = 1$. For any odd p the moment $m_p^{(n)} = 0$ due to the symmetry of Gaussian distribution. Hence, (3) obtains the form

$$m_p^{(g)} = \sum_{k=0}^{[p/2]} \binom{p}{2k} \sigma^{2k} (2k-1)!! \cdot m_{p-2k}^{(f)}. \tag{5}$$

We can see that the moment of the noisy image histogram equals the moment of the clear image histogram plus some additional terms consisting of the moments of h_f of lower orders multiplied by a certain power of σ. For the first few moments we have

$$\begin{aligned} m_1^{(g)} &= m_1^{(f)}, \\ m_2^{(g)} &= m_2^{(f)} + \sigma^2, \\ m_3^{(g)} &= m_3^{(f)} + 3\sigma^2 m_1^{(f)}, \\ m_4^{(g)} &= m_4^{(f)} + 6\sigma^2 m_2^{(f)} + 3\sigma^4, \\ m_5^{(g)} &= m_5^{(f)} + 10\sigma^2 m_3^{(f)} + 15\sigma^4 m_1^{(f)}. \end{aligned}$$

To obtain noise-resistant descriptors, we have to eliminate the parameter σ. This can be done such that the second-order moment is employed to eliminate σ in certain algebraic expressions which contain other moments. In particular, if we define the histogram descriptors (features) as

$$I_p = \sum_{k=0}^{[p/2]} (2k-1)!! \cdot \binom{p}{2k} m_{p-2k} (-m_2)^k. \tag{6}$$

then for any integer $p \geq 0$, the descriptor I_p is fully independent of the image noise regardless of the noise variance. In other words, the I_p value of an arbitrary noisy instance is the same as that of the original, and can be calculated without any denoising or estimating the noise variance (the formal proof of this assertion can be accomplished via induction over p but we skip it due to the limited space).

We use I_p values as histogram features for CBIR. Along with their resistance to noise, they provide "almost complete" representation of the histogram. Having a full sequence of $I_p, \quad p = 1, 2, ...,$ (remember that always $I_0 = 1$) we can recover from (6) all moments of the original histogram except $m_2^{(f)}$. This has a profound reason – since I_p is insensitive to noise, we cannot in principle recover the noise parameter σ, which influences $m_2^{(g)}$. Hence, we could recover the shape of the image histogram while its variance is a free parameter. This also corresponds to the fact that for any image $I_2 = 0$ while all other I_p's are valid. In other words, the full sequence of I_p's provides as much information about the image as its histogram itself with one degree of freedom allowing to incorporate an arbitrary unknown Gaussian smoothing of the histogram. In practice, we of course use only a finite set of these features, the number of which is determined by the user depending on the similarity of the images in the database – the more similar images are to be discriminated, the more histogram features we need. For databases with dissimilar images, only few (typically between 6 and 10) features are sufficient for histogram characterization, which provides an excellent compression ratio.

The intuitive meaning of the I_p's can also be understood as follows. The joint null-space of all I_p's is formed by all Gaussians, so the I_p's define the "distance" between the given histogram and the nearest Gaussian distribution. Equivalently, the I_p's actually measure the non-Gaussian component of the histogram.

It is worth mentioning that all above equations remain valid if we use *central moments* of the histogram instead of the general ones. In that way we achieve the invariance of the method to the overall brightness of the images without any histogram normalization.

IV. Experiments

A. Invariance on simulated AWGN

In the first experiment we demonstrated the invariance property of I_p (6) on pictures with simulated noise. We use testing database of 1000 pictures randomly gathered from Flickr[2]. Average size of the picture is 1.3Mpx and all pictures were converted to grayscale levels.

For each sample picture in the database, we created its noisy version by adding a zero-mean gaussian white noise of various variance. It should be noted that even though the original grayscale image values range from 0 to 255, we do not cut off the values of noisy image so they can range from negative values to values higher than 255.

For each picture and each signal-to-noise ratio (SNR), we extracted two histograms - h_f of the original image and h_g of its noisy version. To show that the invariants I_p give the same results for both clear and noisy picture versions, we calculated the ratio

$$r = \frac{I_p^{(f)}}{I_p^{(g)}}, \tag{7}$$

where we have applied invariant function (6) on histogram of the original image f divided by the invariant applied on histogram of the noisy image g. In Fig. 1 we show the distribution of ratio r for invariants of orders $p = 3, 6, 10$ and 10 different SNRs from the range 5 to 32. It can be clearly seen that majority of the ratios almost equal 1. It is also obvious that the variance of the distribution of r increases as the SNR decreases. The fact that the ratio is not precisely 1 for all cases is because the randomly generated noise is not always exactly Gaussian. Distributions of all three chosen invariant orders are quite similar. However, the higher is the order of the invariant function, the more significant is the influence of numerical errors. This can be observed as a higher variance of distributions in higher-order boxplots. This is an experimental verification that I_p is invariant under additive Gaussian noise.

[2]In all our experiments we use original photographs without any postprocess modifications. Pictures are from the same set that authors of [5] have used.

B. Invariance on real pictures

In the second experiment we demonstrate the invariance of (6) on photographs captured by a compact camera SONY Cyber-shot. This is much more challenging situation namely because of the value cut-offs and the presence of non-Gaussian noise components, which violate the normality of the noise distribution.

We captured 20 different scenes under various light conditions. The light was always low to get a noticeable noise and by light changes we controlled (at least roughly) the noise variance. We toke each scene 20-times and then we estimated the clear image by time-averaging, since under low light it was impossible to obtain the clear image directly (see Fig. 2 for an example). To extract the noise itself, we subtracted the estimated clear image from the noisy version which helps us to estimate the SNR of the captured scenes. We realized that in this experiment the SNR ranges from 25 to 14.

As in the previous experiment, we evaluated the ratio (7) of invariant functions on histograms of noisy and clear pictures. To show the invariance property, the ratio r should be near to 1. Unlike the simulated noise, the real camera noise is subject to cut-off and the histogram support is bounded by values 0-255. This causes the input data for (6) does not meet the required theoretical assumptions perfectly. Anyway, the results of the invariants are quite satisfactory as we can see in Fig. 3. The median of the ratios equals almost 1 for all chosen invariant orders $p = 3, \dots, 10$ and also the majority of invariant ratios are very close to 1. For a comparison and to show that this property is far from being obvious, we calculated the same ratios also for the histogram moments themselves. As one can see in Fig. 3, their behavior is dramatically different and they do not exhibit any invariance to noise.

C. Image Retrieval

Content-based image retrieval is a challenging task where the user selects a query image to retrieve a list of "similar" images (the similarity measure is pre-defined by the user, here we measure the similarity by image histograms) from a large database of pictures. Natural requirement is to avoid mismatches where CBIR method returns images that are not related to the query image. For human perception, two images with the same content seems similar even though one of them is affected by noise. On the other hand, CBIR methods based on comparing image histograms are sensitive to noise that modifies the histogram (see Fig. 5) and therefore standard methods may produce many mismatches. If the database system contains pictures of a similar histogram and either the input query image or the database images are affected by noise, then the danger of mismatches is high.

The aim of this experiment is to show practical application of the proposed invariants (6) to CBIR. In this experiment the database contained clear images (or at least images with invisible noise) while the query image was always a noisy version of one database image. To make the task challenging, we intentionally included pictures of similar histograms into the database. We randomly gathered 71842 photographs from Flickr and clustered them into 314 clusters based on histogram similarity. Thanks to this, we always limited the search to the respective cluster only, which avoids numerous useless comparisons and speeds-up the test without decreasing its significance. It should be noted that each cluster may contain visually similar as well as very different images, see Fig. 4 for an example.

We performed 31400 queries where the query image was always a noisy version of some database image. We created query images of five SNR levels (5, 10, 15, 20 and 25). For each SNR we generated 20 different instances of the noise. The histograms of query images were heavily smoothed due to the noise (see Fig. 5). We compared four retrieval methods, all of them based on the histograms. In the first method we used the invariants $I_3, ..., I_{10}$ as in the previous experiments. The second method uses plain histogram moments $m_1, ..., m_{10}$. Since I_p and m_p grows rapidly as the order p increases, we normalized both I_p's and m_p's to keep them in a comparable range. In the third method we used the complete histograms (256 bins) as feature vectors and matched them by measuring their ℓ_2 distance. In the last method, we denoised the query images first (we applied a wavelet-based denoising [1]) and then applied full histogram matching as before.

Since we know the ground truth, we can evaluate the correct retrieval rate. Fig. 6 shows the success rate of the four methods as a function of the SNR. The results of the first three methods confirmed our theoretical expectation. The invariant-based retrieval yields the best results, complete histogram is the worst choice and the performance of plain moments is somewhere in between. The differences between the methods increases as the SNR decreases. At the same time, the success rate of each method also increases with an increasing SNR. However, this drop-off is much more severe in case of the complete histogram than in case of the invariants. The fact that the plain moments perform better than a complete histogram might look a bit surprising at the first sight. The explanation is that we used only 10 low-order moments that describe global characteristics of the histogram which are less influenced by the noise than the complete histogram itself. The result of the fourth method – denoising followed by histogram matching – is worse than we originally expected. We assumed the method should perform comparably to the invariants but actually it is much worse and it is even worse than the plain moments. The reason is that the denoising decreases the noise level in the image but does not restore the original histogram well. Another drawback of this approach is that it requires a significant extra time to perform the denoising. We tried to replace the wavelet denoising by BM3D algorithm [2], which is one of the best ranked existing denoising methods and re-run the experiment. However, the BM3D algorithm is so slow (10 minutes for one query image with 20 instances of noise) that we run it on a small subset only with the conclusion that the success rate is comparable to that of wavelet denoising. Hence, an interesting conclusion is that denoising followed by histogram matching is absolutely not suitable in terms of both success rate and speed, regardless of the particular denoising algorithm.

V. Conclusion

Histogram of a noisy image, both visual appearance and common numerical characteristics, use to be significantly affected by additive noise in the image. Provided the noise is Gaussian, we proposed novel histogram descriptors which are invariant w.r.t. noise. We proved that along with the theoretical invariance the descriptors are sufficiently robust on real images corrupted by thermal, electronic, and shot noise. As demonstrated experimentally, the proposed descriptors can be used as the features in CBIR if the database and/or query images are heavily noisy and standard descriptors fail. We verified that this approach is more efficient in terms of retrieval rate and speed than image denoising followed by histogram matching.

Although this paper has dealt with graylevel images only, the method can be applied to color images as well. In such a case we have to work with three single histograms or with 3D histograms. In the latter case, we can employ a 3D analogy to Eq. (6). Such modification is computationally more expensive and brings certain problems introduced by histogram sparsity but, assuming that the noise instances in individual color bands are mutually uncorrelated, performs rather straightforward generalization of the graylevel case.

TABLE I. Image Database Summary

Number of databases:	314
Total number of pictures:	71842
Number of queries:	31400
Average pictures count per DB:	229

Acknowledgement

This work has been supported by the Czech Science Foundation under the Grant No. GA13-29225S.

References

[1] R.R. Coifman and D.L. Donoho. Translation-invariant de-noising. In Anestis Antoniadis and Georges Oppenheim, editors, *Wavelets and Statistics*, volume 103 of *Lecture Notes in Statistics*, pages 125–150. Springer New York, 1995.

[2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *Image Processing, IEEE Transactions on*, 16(8):2080–2095, Aug 2007.

[3] Sangoh Jeong. Histogram-based color image retrieval. Technical report, Stanford University, 2001. Psych221/EE362 Project Report.

[4] Guang-Hai Liu, Lei Zhang, Ying-Kun Hou, Zuo-Yong Li, and Jing-Yu Yang. Image retrieval based on multi-texton histogram. *Pattern Recogn.*, 43(7):2380–2389, July 2010.

[5] B. Mahdian, R. Nedbal, and S. Saic. Blind verification of digital image originality: A statistical approach. *Information Forensics and Security, IEEE Transactions on*, 8(9):1531–1540, 2013.

[6] M.K. Mandal, T. Aboulnasr, and S. Panchanathan. Image indexing using moments and wavelets. *IEEE Transactions on Consumer Electronics*, 42:557–565, 1996.

[7] G. Pass and R. Zabih. Histogram refinement for content-based image retrieval. In *Applications of Computer Vision, 1996. WACV '96., Proceedings 3rd IEEE Workshop on*, pages 96–102, 1996.

[8] Michael J. Swain and Dana H. Ballard. Color indexing. *Int. J. Comput. Vision*, 7(1):11–32, November 1991.

[9] Wang Xiaoling. A novel circular ring histogram for content-based image retrieval. In *Education Technology and Computer Science, 2009. ETCS '09. First International Workshop on*, volume 2, pages 785–788, 2009.

[10] P.-T. Yap and R. Paramesran. Content-based image retrieval using legendre chromaticity distribution moments. *Vision, Image and Signal Processing, IEE Proceedings* -, 153(1):17–24, 2006.

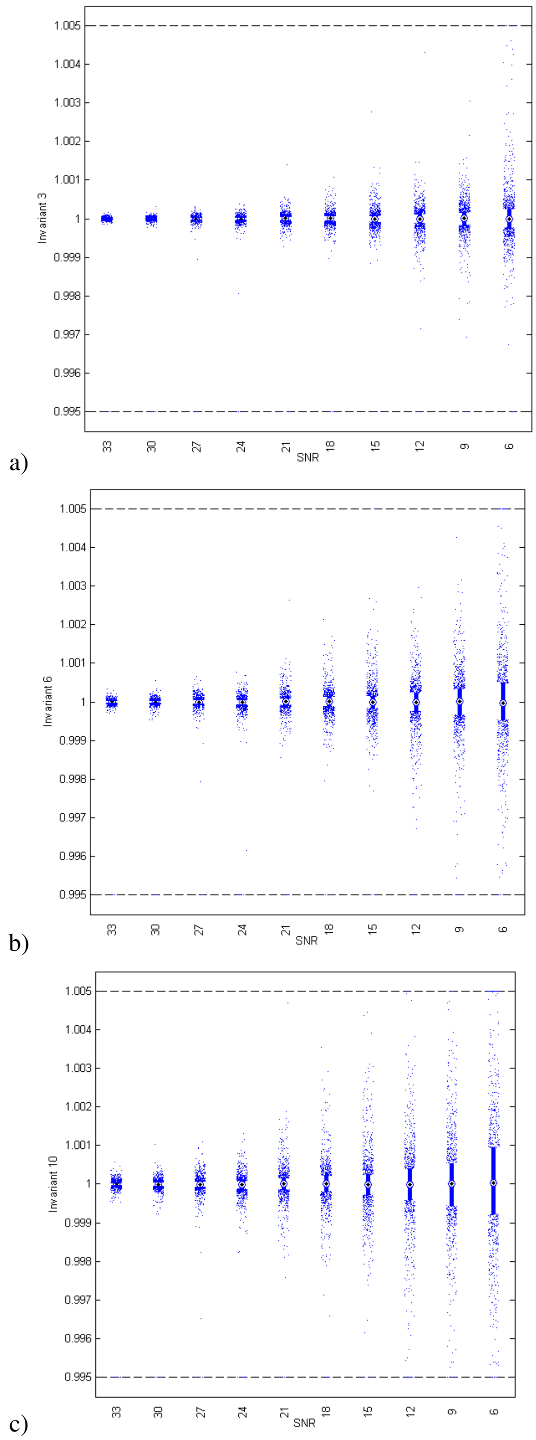


Fig. 1. The boxplots show the distribution of 1000 ratios of invariants calculated on original images and their noisy versions. The boxplots from top to bottom show the results for invariant orders 3, 6 and 10 respectively. The central mark shows the median, thick bar depicts 50% of the data between 25^{th} and 75^{th} percentiles. Outliers outside this range are marked as dots.

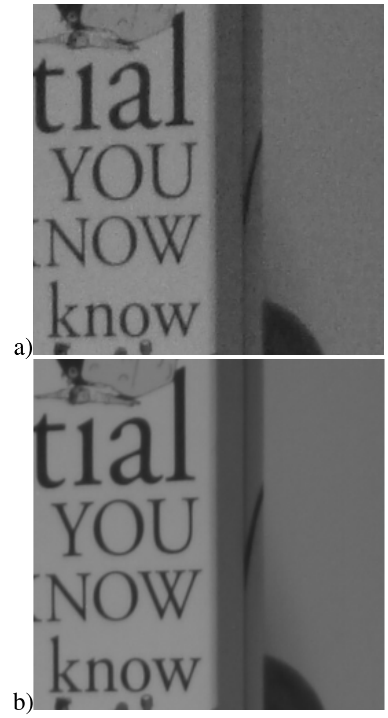


Fig. 2. A crop of the scene photographed in low light. Originally captured noisy image (a) and the noise-free image constructed by averaging 20 noisy frames of the same scene (b).

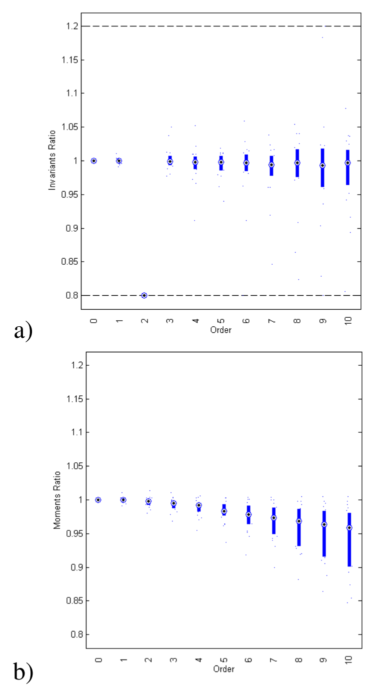
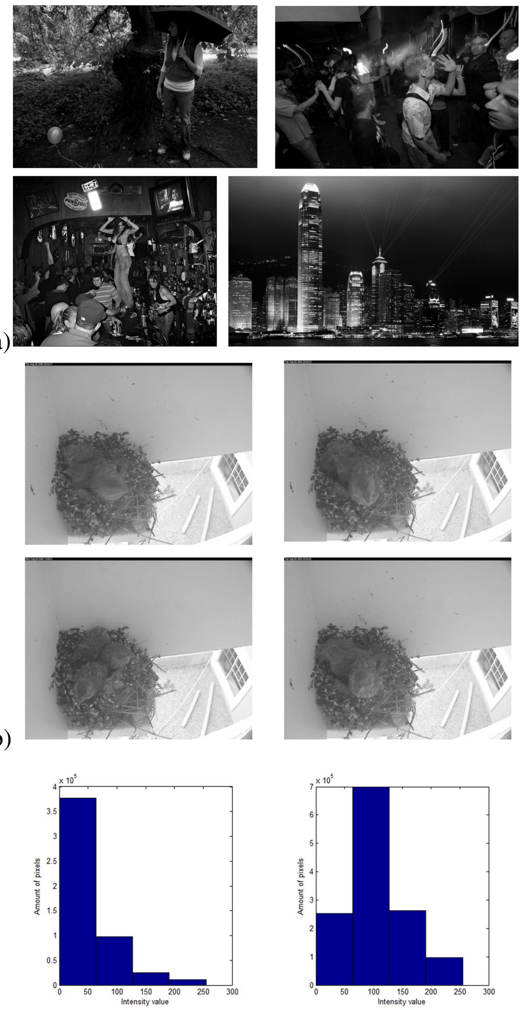


Fig. 3. a) The boxplots show the ratio (7) of invariants calculated on histograms of real clear and noisy images. Central mark is the median of the distribution. Thick bar depicts 50% of the data between 25^{th} and 75^{th} percentiles. Outliers outside this range are marked as dots. b) The boxplots show the same ratio where plain moments m_p were used instead of invariants I_p. The invariance is heavily violated.



a)

b)

c)

Fig. 4. Sample images from the test database. Pictures were clustered according to their histogram similarity. When considering histograms simplified into four bins, all pictures within one cluster have the same simplified histogram. In a) and b) there are previews of pictures from two clusters with corresponding histograms in c) (on the left is a histogram for cluster a) and on the right for cluster b)). Some clusters contain pictures that have the same histogram but look differently (e.g. cluster a)), some clusters contains pictures that even look very similarly (e.g. cluster b)).

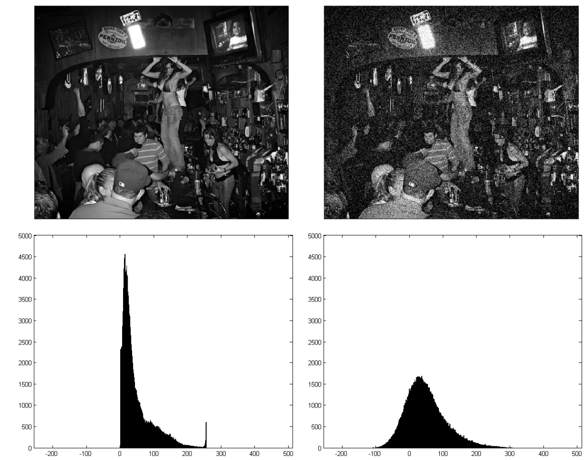


Fig. 5. Example of the query image affected by noise (SNR=5) (right) and the clear version of the image in the database (left). In the bottom there are histograms of the images. It can be seen that the noise causes significant modification of the histogram.

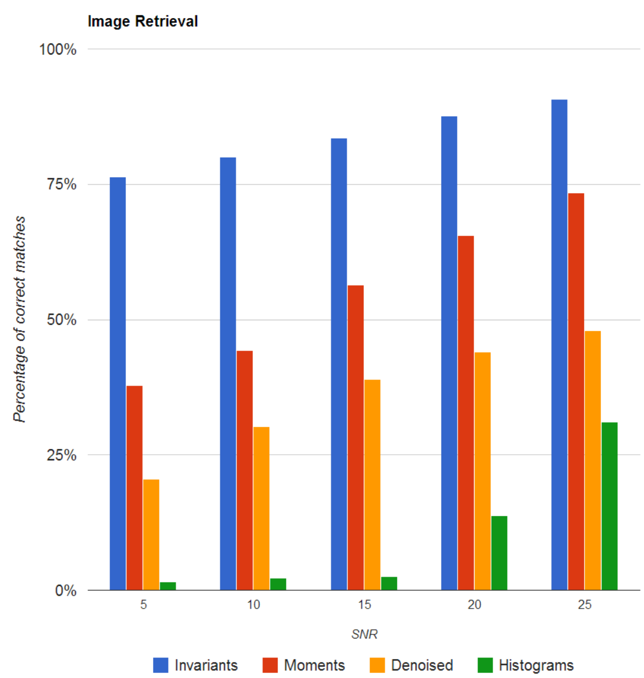


Fig. 6. Image retrieval experiment results. The graph shows percentages of correct matches for 6280 queries for each SNR (total is 31400 queries in 314 databases). The method based on invariants outperforms the moments and the complete histogram matching of original and denoised images.