Pattern Recognition Letters 69 (2016) 72–81

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Robust histogram-based image retrieval☆

Cyril Höschl IV, Jan Flusser*

Institute of Information Theory and Automation, Czech Academy of Sciences, Pod vodárenskou věží 4, 182 08 Praha 8, Czech Republic

ARTICLE INFO

Article history:
Received 27 March 2015
Available online 12 November 2015

Keywords:
Image retrieval
Noisy image
Histogram
Convolution
Moments
Invariants

ABSTRACT

We present a histogram-based image retrieval method which is designed specifically for noisy query images. The images are retrieved according to histogram similarity. To reach high robustness to noise, the histograms are described by newly proposed features which are insensitive to a Gaussian additive noise in the original images. The advantage of the new method is proved theoretically and demonstrated experimentally on real data.

1. Introduction

Since the appearance of the first image databases in the 80's, image retrieval has been the goal of intensive research. Early methods did not search the images themselves but utilized some kind of metadata and image annotation (tagging) to retrieve the desired images. As many large-scale databases do not contain any annotations (manual annotation is expensive and laborious while automatic tagging is still under development), content-based image retrieval (CBIR) methods have become 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 predefined metric) to a given query image. CBIR methods do not rely on a 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 a majority of users and the users feel this similarity at the first sight without a detailed exploration of the image content. This requirement, along with the need for a fast system response, has led 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 the 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 from each other. On the other hand, those images that have substantially different histograms are rarely rated by observers as similar. Another attractive property of the histogram is that, if normalized to the image size, it does not depend on image translation, rotation and scaling, 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 of the contrast and brightness of the image. Hence, the histogram established itself as a meaningful image characteristic for CBIR [7–9].

The 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 it is also redundantly detailed. It is sufficient and computationally efficient to capture only the prominent features of the histogram and suppress the insignificant details. To do so, some authors compressed the histogram from the full color range into few bins [3,4] while some others represented 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 a high compression on one hand and an accurate representation on the other hand. It is very natural to get inspired by a clear analogy between histogram of an image and a probability density function (pdf) of a random variable. In probability theory, the pdf is usually

☆ This paper has been recommended for acceptance by Nappi Michele.

* Corresponding author. Tel.: +42 2 6605 2357; Fax: +42 2 6641 4903.
E-mail addresses: hoschl@utia.cas.cz (C. Höschl IV), flusser@utia.cas.cz (J. Flusser).

http://dx.doi.org/10.1016/j.patrec.2015.10.012

characterized by its moments, so it is worth applying the same approach in the histogram-based CBIR [6,10].

The 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 a 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 at the expense of fine image details.

Although the noise in digital photographs is an issue we can neither avoid nor ignore, very little attention has been paid to developing noise-resistant CBIR methods. The authors of the papers on CBIR have either skipped this problem altogether or rely on denoising algorithms applied to all images before they enter the database. Such a solution, however, is not convenient or even not realistic, because the denoising inevitably introduces artifacts such as high-frequency cutoff, requires additional time, and mostly also needs a cooperation of the user in tuning the 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 Gaussian noise. The core idea of the method is a proper representation of the histogram by certain characteristics, which are not affected by the noise. We stress that the paper does *not* aim to evaluate in which tasks and for what purposes a histogram-based CBIR is appropriate. We rather show how, if it is appropriate, it should be implemented in the case of noisy database and/or noisy query images. Our method does not perform any denoising and cannot replace it in the applications where the noise should be suppressed to improve the visual quality of the image.

In the rest of the paper, we first describe the noise model we are working with and show how this noise influences the image histogram. Then we present a noise-resistant representation of the histogram and demonstrate the advantage of this representation in CBIR. In the experimental part, we compare the new method with several traditional approaches and demonstrate their advantages on a database of more than 70,000 images and 30,000 queries.

2. The noise model

As we already mentioned, we primarily consider the thermal noise and electronic background noise of consumer cameras. It is a common belief that such noise n can be modeled as a 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. If this assumption were true, the noise normalized histogram h_n would have 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 $g = f + n$ would then be a convolution of the original histogram and the noise histogram

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

Apparently, such an ideal model can hardly be encountered in practice. Let us however demonstrate on an example that it performs a reasonable approximation of a real noise. In Fig. 4(a), we can see a clip of size 427 × 386 pixels of a real noisy image taken under low-light conditions. In order to separate f and n, we took this image repeatedly twenty-times and we estimated f by time-averaging these 20 frames (see Fig. 4(b)). This allows us to calculate all three histograms h_g, h_f, and h_n and a synthetic histogram $h_c = h_f * h_n$ (see Fig. 1 from top to bottom). We can see that the noisy picture histogram in Fig. 1(c) matches the synthetic histogram in Fig. 1(d). Additionally, in Fig. 2 we can see the normality plot of the image noise n is very close to a normal distribution. We repeated this experiment for many images with the same conclusion. Hence, we consider our noise model acceptable and use it for deriving a proper histogram representation.

3. Histogram representation resistant to image noise

In this section, we present a representation of the image histogram by descriptors which are not affected by AGWN. These descriptors are based on the statistical moments of the histogram, which is a common approach to the characterization of pdf's in probability theory. 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 $p = 0, 1, 2, \ldots,$ is called *general moment* of the pdf. Clearly, $m_0 = 1$, m_1 equals the mean value and m_2 would equal the variance (if the histogram was centralized) of X. In general, the existence (finiteness) 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.[1] In this sense moments provide a complete and non-redundant description of a pdf/histogram.

Unfortunately, the histogram moments 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

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

This assertion can easily be proved just using the definitions of moments and of convolution. Since the noise is supposed to be 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, $k!! = 1 \cdot 3 \cdot 5 \ldots k$ for odd k, and by definition $(-1)!! = 0!! = 1$. For any odd p the moment $m_p^{(n)} = 0$ due to the symmetry of the 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

$$m_1^{(g)} = m_1^{(f)},$$

$$m_2^{(g)} = m_2^{(f)} + \sigma^2,$$

$$m_3^{(g)} = m_3^{(f)} + 3\sigma^2 m_1^{(f)},$$

[1] A more general moment problem is well known from theory of probability: can a given sequence be a set of moments of some compactly-supported function? The answer is yes if the sequence is completely monotonic.