Image Retrieval by Local Evaluation of Nonlinear Kernel Functions around Salient Points

Alaa Halawani and Hans Burkhardt
Albert-Ludwigs-University of Freiburg
Chair of Pattern Recognition and Image Processing
79110 Freiburg, Germany
{halawani|burkhardt}@informatik.uni-freiburg.de

Abstract

Feature histograms based on the evaluation of Haar integrals with nonlinear kernel functions were used successfully for the purpose of invariant content based image retrieval. In addition to being invariant to rotation and translation, the features have the advantage of preserving structural information of the image. The work presented here concentrates on the idea of calculating these features by evaluating the kernel functions around a small set of preselected points. These points are called the salient points and represent, together with their neighborhood, the most important visual information in an image. The use of these salient points leads to a better representation of the image. Compared to previous work, experiments show that this method gives better retrieval results without introducing extra computational overhead.

1. Introduction

Content-Based Image Retrieval (CBIR) has gained more and more attention in the last few years. As the size of the digital image database gets larger, the manual search for similar images in this database becomes a tedious work. The main aim of CBIR is to search for similar images in a given database based on an expressive representation of the most important information held in these images. The process of finding these expressive information is known as "Feature Extraction". The assumption is that similar images should have similar representations (features). Despite all the advances in the field of CBIR, the task of finding good features that adequately represent an image is still a challenging task.

Recently, Siggelkow et al. [6, 9, 5] have used rotation- and translation-invariant color and texture feature histograms for image retrieval. These features are based on the invariant integration described in [4]. Unlike the ordinary histogram, the invariant integral features have the advantage of capturing the local structure held in the image [6]. Experimental results have shown that these features demonstrate a very good capability in retrieving images. However, the main disadvantage is that the computation of the invariant features over the whole image is time consuming. In order to reduce the computation complexity, Siggelkow and Schael [8] have estimated the invariant features using the Monte-Carlo method. They compute the features for a set of randomly generated points and directions.

In this paper we aim at enhancing the performance of such a retrieval system by concentrating on extracting these features from areas of high relevance in the image under consideration. The features extracted from these patches should give a more discriminant representation of the image, the matter that will lead to better retrieval results. These areas can be charecterised by a small set of image points, called salient points, together with their neighborhood. In order to determine these points, we use the salient point extraction algorithm proposed by Loupias et al. in [1]. Because feature computation is limited to a small set of pixels, there will be no increase in the computation complexity. Additionally, the use of the salient points is expected to show more robustness to situations where objects are scaled or presented in different views.

The paper is organized as follows: In section 2 we explain the process of calculating the invariant features. Section 3 describes the algorithm used for salient point extraction. A summary of the experimental results is presented in section 4. Finally, a conclusion is given in section 5.

2. Invariant color and texture features

Following is a brief description of the calculation of the rotation- and translation-invariant features. Details can be found in [4].

The idea of constructing invariant features is to apply a nonlinear kernel function $f(\mathbf{I})$ to the gray-valued image, $\mathbf{I}$, and to integrate the result over all possible rotations and translations (Haar integral over the Euclidean motion); i.e.,

$$IF(\mathbf{I}) = \frac{1}{2\pi MN} \int_{r=0}^{M} \int_{c=0}^{N} \int_{\theta=0}^{2\pi} f(g(r,c,\theta)\mathbf{I}) d\theta dr dc \tag{1}$$

where $IF(\mathbf{I})$ is the invariant feature of image, M, N are the dimensions of the image, and g is an element in the transformation group G (which consists here of rotations and translations).

Because of the discrete nature of the image, IF is approximated by choosing r and c to be integers and by varying θ in a discrete manner producing q samples:

$$IF(\mathbf{I}) \approx \frac{1}{qMN} \sum_{r=0}^{M-1} \sum_{c=0}^{N-1} \sum_{j=0}^{q-1} f(g(r,c,\theta = j\frac{2\pi}{q})\mathbf{I}) \tag{2}$$

Bilinear interpolation is applied when the samples do not fall onto the image grid.

The above equation suggests that invariant features are computed by applying a nonlinear function, f, on the neighborhood of each pixel in the image, then summing up all the results to get a single value representing the invariant feature. Using several different functions finally builds up a feature space.

Much of the local information is lost by summing up the local results. This makes the discrimination capability of the features very weak. In order to preserve the local information, Siggelkow et al. [6, 9, 5] replaced the summation ($\sum_r \sum_c$) by histogramming.

It is also possible to replace all the summations by a histogram operation [7], i.e.,

$$IF(\mathbf{I}) = \text{hist}\left(\left\{ f(g(r,c,\theta = j\frac{2\pi}{q})\mathbf{I}) \;\middle|\; \begin{array}{l} r = 0, \cdots, M-1, \\ c = 0, \cdots, N-1, \\ j = 0, \cdots, q-1 \end{array} \right\} \right) \tag{3}$$

Invariant features can be either color or texture features, depending on the chosen kernel function. Invariant color features can be computed by applying the so-called "*monomial kernels*" which have the form:

$$f(\mathbf{I}) = \left(\prod_{p=0}^{P-1} \mathbf{I}(x_p, y_p) \right)^{\frac{1}{P}} \tag{4}$$

In order to construct texture features, a "*relational kernel*" function [3] is to be applied. This kernel has the form:

$$f(\mathbf{I}) = rel(\mathbf{I}(x_1, y_1) - \mathbf{I}(x_2, y_2)) \tag{5}$$

where

$$rel(\gamma) = \begin{cases} 1 & \text{if } \gamma < -\epsilon \\ \frac{\epsilon - \gamma}{2\epsilon} & \text{if } -\epsilon \leq \gamma \leq \epsilon \\ 0 & \text{if } \epsilon < \gamma \end{cases} \tag{6}$$

This kind of kernels was introduced in [3] and is based on the Local Binary Pattern (LBP) texture features [2], which map the relation between a center pixel and its neighborhood pixels into a binary pattern. Equation 6 extends the LBP operator to give values that fall in $[0, 1]$. This is done in order to get rid of the discontinuity of the LBP operator which makes this feature sensitive to noise.

3. Salient points from wavelets

The salient point extraction algorithm presented here was introduced by Loupias and Sebe [1]. The assumption is that image points, where high variations occur, represent important information in the image (areas of high relevance) and are extracted. One can study the variations that are present in an image using the wavelet analysis which allows for multiresolution representation of a signal (image). The algorithm starts from the coarsest resolution after representing the image in the wavelet domain. Taking the absolute value of the wavelet coefficients, a high coefficient c_l at a coarse resolution corresponding to level l of the wavelet transform, must have come from an image region with high variations. This coefficient is actually computed from a set of known image points, P. Going backwards to a finer resolution at level $l-1$, one can find a set of wavelet coefficients, $\mathbf{WC}_{l-1}$, that are computed from the same set of image points for the coefficient c_l at level l. These coefficients are called the *children* of c_l [1]. Again, the coefficients in $\mathbf{WC}_{l-1}$ represent the variations of points in P at level $l-1$. The maximum coefficient, $c_{l-1} \in \mathbf{WC}_{l-1}$, represents the highest variation and therefore must have been computed from the most salient subset, $p \subset P$, of the set of image points, P. This coefficient, c_{l-1}, is taken into consideration. We go back again to a finer resolution to investigate the children of c_{l-1}. This procedure is applied recursively until we end up with picking one coefficient at level 1 (level 0 represents the original image). This coefficient represents a number of points in the original image. Among these points, the point with the maximum gradient is chosen and is given a value representing its saliency. This saliency value is equal to the sum of the absolute value of the wavelet coefficients along the whole track:

$$s = \sum_{i=1}^{l} |c_i| \tag{7}$$

Fig. 1 illustrates the process of tracking the wavelet coefficients.

The above scenario is repeated for every wavelet coefficient that exceeds a certain threshold, τ, so that we do not

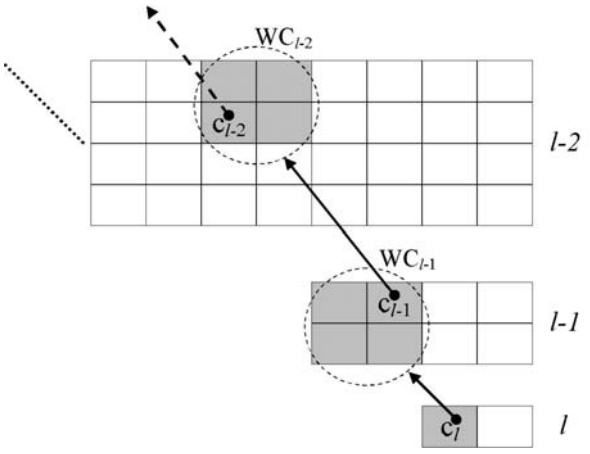


Figure 1. Tracking wavelet coefficients to extract salient points

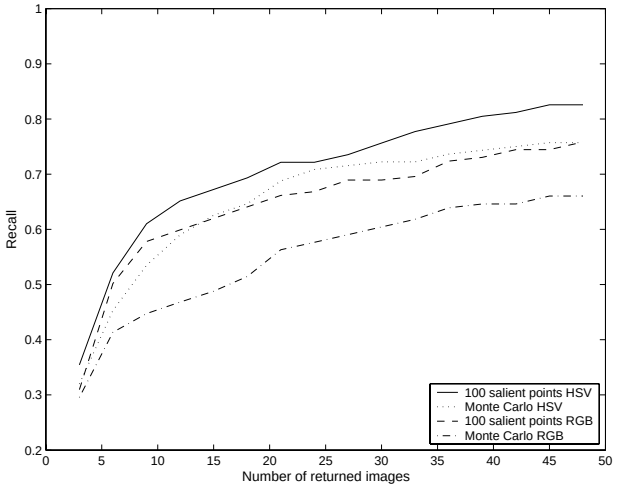


Figure 2. Average recall of 19 query images as a function of images returned to the user

waste time investigating small wavelet coefficients. We end up with a matrix (which we call the *"Saliency map"* here) representing the saliencies of the image pixels. The saliency map is thresholded to end up with a set of chosen saliency points.

4. Results

This section sums up and compares the results that we got by evaluating the invariant features around the extracted salient points with those of the Monte-Carlo method.

4.1. System setup

We conducted our tests on the same database used in [9], which consists of 2343 colored images with a resolution of 384×256.

We use both RGB and HSV color spaces. The salient point extraction algorithm using the Haar wavelet transform is applied to the V Channel of the images in the case of HSV color space or to a gray value copy of the image in the case of RGB space. The threshold τ is set to 0.1. The resulting saliency map is thresholded to give 100 salient points.

Both color and texture features are evaluated on these points varying θ in a step of $\frac{\pi}{24}$ producing 48 samples.

For the texture features, we applied the kernel $f(\mathbf{I}) = rel(\mathbf{I}(3,0) - \mathbf{I}(0,6))$ with $\epsilon = 0.098$ in Equation 6 (image pixel values $\in [0,1]$). The monomial kernel, $f(\mathbf{I}) = (\mathbf{I}(3,0).\mathbf{I}(0,6))^{\frac{1}{2}}$, was chosen to construct the color features. The three channels of both color spaces were taken into consideration when calculating the features.

Two $4 \times 4 \times 4$ Histograms (color Histogram, H_c, and texture histogram, H_t) were constructed.

The features for all images are extracted offline and saved with links to the images in a feature database. The system works by the query-by-example (QBE) methodology. A query image is introduced and its features are computed online and compared with the features of all other images in the database. To compare the histograms of the query image and the database images, we have used the χ^2 measure, which gives an indication of the difference (d) between two histograms:

$$\chi^2(h_q, h_d) = d = \sum_i \frac{(h_q(i) - h_d(i))^2}{h_q(i) + h_d(i)} \qquad (8)$$

where h_q and h_d are the histograms of the query image and an image in the database respectively.

To increase the accuracy of the retrieval, one should integrate the results of both comparisons of texture and color features. Let d_c equal the difference between the query image and a database image based on color, and d_t equal to the difference based on texture; the difference based on both color and texture is given by:

$$d_{total} = \alpha d_c + \beta d_t, \qquad \alpha + \beta = 1 \qquad (9)$$

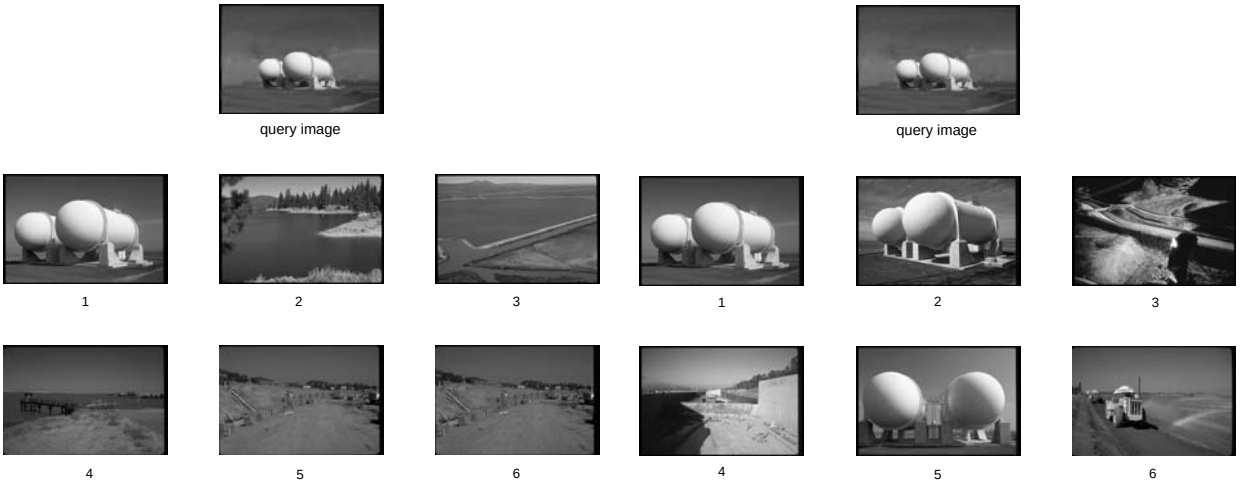
where α and β are weights assigned to the color-based difference and texture-based difference respectively.

4.2. Retrieval results

From the image database we have chosen a set of 19 representative images to conduct our experiments.

(a) Monte-Carlo method (b) 100 salient points

Figure 3. Results of a sample query



(a) Monte-Carlo method (b) 100 salient points

Figure 4. Query in the case of scaling

Figure 5. White dots represent the 100 salient points of images 4 and 5 in Fig. 3(b)

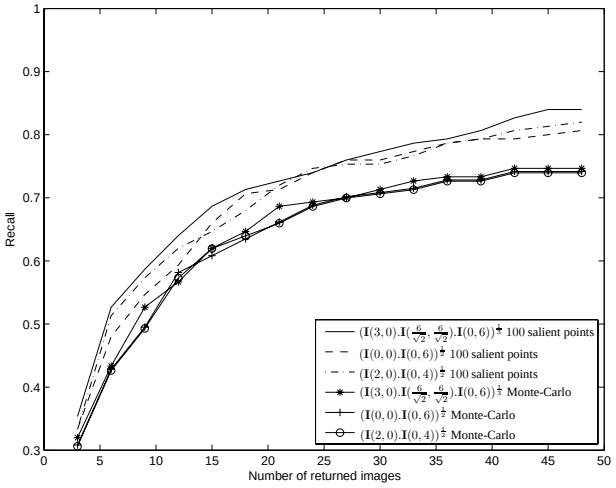


Figure 6. Average recall for the HSV color space using different kernel functions

Setting equal weights for color and texture ($\alpha = \beta = 0.5$), we have tested the performance of the method with 100 salient points against the Monte-Carlo method with 38416 random triples of r, c, and θ.

The query time of both methods is comparable. On average, a query for the 100 salient points lasts $1s$ and $1.3s$ for the Monte-Carlo method.

Fig. 2 shows the average recall of the queries as a function of retrieved images. The recall is defined as:

$$\text{Recall} = \frac{\text{\# of relevant images retrieved}}{\text{total \# of relevant images}} \qquad (10)$$

From the figure, one can observe that the results obtained using the HSV color space are better than those of the RGB space. This observation is expected because of the fact that the HSV color space is approximately perceptually uniform compared to the RGB space. This approximate perceptual uniformity makes the HSV space more suitable for quantization than the RGB space.

It can also be observed that the salient point method performs better in both spaces than the Monte-Carlo method. Furthermore, the salient point method's performance in the RGB space is comparable with the performance of the Monte-Carlo method in the HSV space, and even outperforms it when the number of returned images is small (< 15 images). This behavior is expected because the salient points and their neighborhood represent the most important visual information of the image [1].

As an example, consider the results shown in Fig. 3(a) and 3(b), which show the results of a query using the Monte-Carlo method and 100 salient points, respectively. In Fig. 3(a), only the first two images are considered to be relevant to the query image, while in Fig. 3(b) the first five results are relevant. The images labeled 3, 4, and 5 in Fig. 3(b) represent the same object that is in the query image (the castle) but taken from different views (the query image is a close shot of a part of the castle, images 3 and 4 show the whole castle from different angles, and in image 5 the castle appears with some surrounding). Because retrieval using salient points concentrates only on the most important visual details in the image, most of the points in images 3 and 4 in Fig. 3(b) are concentrated around and in the main object (the castle), the matter that makes their representation close to that of the query image. The same applies for image 5 taking into account that some of the salient points are distributed in the surrounding of the castle, which causes its representation to be further away from the query. Fig. 5 shows images 4 and 5 of Fig. 3(b) with their 100 most salient points represented by the white dots.

Another situation is shown in Fig. 4. The query image is a scaled version of an object that can be found in four of the images of the database. Retrieval results using the Monte-Carlo method show that only one relevant image is retrieved (recall = 0.25), while three relevant results were returned using 100 salient points (recall = 0.75). This shows that the use of the salient points for image retrieval tolerates the situations in which there is scaling better than the Monte-Carlo method.

Further experiments were conducted using other kernel functions than those used above. In all the experiments, the salient point method outperformed the Monte-Carlo method. The results of some of these experiments are shown in Fig. 6. The legend in the figure indicates the monomial

kernels used. Similar relational kernels were used except for the case of the kernel $(\mathbf{I}(3,0).\mathbf{I}(\frac{6}{\sqrt{2}},\frac{6}{\sqrt{2}}).\mathbf{I}(0,6))^{\frac{1}{3}}$ in which we used the same relational kernel as in the above discussion.

5. Conclusion

In this paper we have compared the evaluation of the nonlinear kernel functions around salient points with the evaluation of the same functions using the Monte-Carlo method for the purpose of image retrieval. Both color and texture features were used. Several experiments with different kernel functions were carried out. In all of the experiments, the salient point method gave better results than the Monte-Carlo method. This is due to the fact that the method of salient points extracts the features from the most important parts of the image. As was shown, this fact also makes the salient point method less sensitive to object scaling and view changes. Better performance can be achieved when using the HSV color space instead of the RGB color space. This is because the HSV space is approximately perceptually uniform, which leads to better quantization. Very good results were achieved using only 100 salient points.

Acknowledgement

Alaa Halawani would like to thank the German Academic Exchange Service (DAAD) for granting him a scholarship for his PhD studies at the University of Freiburg in Germany.

References

[1] E. Loupias and N. Sebe. Wavelet-based Salient Points for Image Retrieval. Technical Report TR. 99.11, Laboratoire Reconnaissance de Formes et Vision, INSA Lyon, 1999.

[2] T. Ojala, M. Pietikäinen, and T. Mäenpää. Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns. In *Proc. Sixth European Conference on Computer Vision*, pages 404–420, Dublin, Ireland, 2000.

[3] M. Schael. Invariant Grey Scale Features for Texture Analysis Based on Group Averaging with Relational Kernel Functions. Technical Report 1/01, Albert-Ludwigs-Universität, Freiburg, Institut für Informatik, January 2001.

[4] H. Schulz-Mirbach. Invariant Features for Gray Scale Images. In G. Sagerer, S. Posch, and F. Kummert, editors, *17th DAGM - Symposium "Mustererkennung"*, pages 1–14, Bielefeld, 1995. Springer.

[5] S. Siggelkow. *Feature Historgrams for Content-Based Image Retrieval*. PhD thesis, Albert-Ludwigs-Universität, Freiburg, December 2002.

[6] S. Siggelkow and H. Burkhardt. Image Retrieval Based on Local Invariant Features. In *Proceedings of the IASTED International Conference on Signal and Image Processing (SIP)*, pages 369–373, Las Vegas, Nevada, USA, 1998.

[7] S. Siggelkow and H. Burkhardt. Improvement of Histogram-Based Image Retrieval and Classification. In *In Proceedings of the International Conference on Pattern Recognition*, volume 3, pages 367–370, Quebec, Canada, September 2002.

[8] S. Siggelkow and M. Schael. Fast Estimation of Invariant Features. In W. Förstner, J. M. Buhmann, A. Faber, and P. Faber, editors, *Mustererkennung, DAGM 1999, Informatik aktuell*, pages 181–188, Bonn, September 1999.

[9] S. Siggelkow, M. Schael, and H. Burkhardt. SIMBA - Search IMages By Appearance. In B. Radig and S. Florczyk, editors, *Proceedings of 23rd DAGM Symposium, number 2191 in LNCS Pattern Recognition*, pages 9–16. Springer, September 2001.