Combining Interest Points and Edges for Content-based Image Retrieval *

Junqiu Wang and Hongbin Zha
National Laboratory on Machine Perception
Peking University
Beijing 100871, China
jerywang@public3.bta.net.cn, zha@cis.pku.edu.cn

Roberto Cipolla
Department of Engineering
University of Cambridge
Cambridge, CB2 1PZ, UK
cipolla@eng.cam.ac.uk

Abstract— **This paper presents a novel approach using combined features to retrieve images containing specific objects, scenes or buildings. The content of an image is characterized by two kinds of features: Harris-Laplace interest points described by the SIFT descriptor and edges described by the edge color histogram. Edges and corners contain the maximal amount of information necessary for image retrieval. The feature detection in this work is an integrated process: edges are detected directly based on the Harris function; Harris interest points are detected at several scales and Harris-Laplace interest points are found using the Laplace function. The combination of edges and interest points brings efficient feature detection and high recognition ratio to the image retrieval system. Experimental results show this system has good performance.**

I. Introduction and related work

Content-Based Image Retrieval (CBIR) aims to find images with the required characteristics in an image database. A successful CBIR system is based on distinctive feature detection, effective feature representation, and efficient indexing. In recent years, interest point detectors and descriptors are employed in many CBIR systems [5], [10], [11]. Representation based on local features is robust against background changes and occlusion because of the small sizes of local features. However, local features are not always enough for discrimination: It is difficult to differentiate Fig.1.(a) from Fig.1.(b) if color information is not considered. No reliable interest points can be detected in Fig.1.(c) and Fig.1.(d). These examples show that color information is very important for image retrieval.

Color is one of the most intuitive features for visual recognition. The color histogram that describes the statistics of colors in an image is widely used in CBIR systems. It is easy to compute and is insensitive to small changes in viewing positions and partial occlusion. However, the color histogram has several disadvantages: firstly, the color histogram overlooks salient local features that have discriminative ability. Secondly, if the background of the image has the dominant color, the representation is not effective when the background changes. Thirdly, the traditional color histogram does not include any spatial information. Images with similar color distributions may have very different appearance.

In this work, we aim to combine the SIFT descriptor with color histogram to describe an image. In [1], an image is decomposed into three kinds of regions: smooth areas, edges and corners. There are little signal changes in a smooth area, whereas edges and corners contain the maximal amount of information necessary for a successful recognition. The image is well represented if corners and edges are detected and characterized efficiently. In this work, the detection of edges and corners is achieved using the Harris function [1]. The representation of edges is based on the color histogram. The changes of color in an image occur at the color edges, therefore the color distribution on the pixels around edges is similar to that of the image [12]. The edge color histogram partly solves the problem caused by background changes. Shim and Choi propose the edge color histogram computed only on edges [12]. However, the detection of edges in [12] is a little complex and they do not utilize information of interest points. Mikolajczyk and Schmid extend the Harris detector to the Harris-Laplace detector which detects Harris interest points at several scales and then selects the right scale by finding the maximum of the Laplace function [5]. Harris-Laplace detector is able to deal with the problem of scale changes. In this work, the Harris-Laplace interest points are described by the SIFT descriptor that is better than other descriptors [6].

Several CBIR systems incorporate spatial information about colors to improve upon the histogram method. In [12], the color pixels on the edges are classified into three kinds: horizontal, vertical and non-orientation. In this work, color pixels are projected on the horizontal and the vertical directions to include spatial information. Using this approach, one color pixel may contribute to both directions. Thus the spatial description is more accurate than [12].

The interest points detected in an image are described by the SIFT descriptor which is a 128-dimensional vector. Image retrieval based on directly nearest neighbor searching is not efficient enough [15]. The *Vector Space Model*(VSM), which has been successfully used in text retrieval, is employed in this work to accelerate the localization process. In the VSM, a collection of documents is represented by an inverted index. In the index, each document is a vector and each dimension of the vector represents a count of the occurrences for a word [15]. The documents for

*This work was sponsored in part by the National Science Foundation of China under grant No. 6017525.

0-7803-9134-9/05/$20.00 ©2005 IEEE

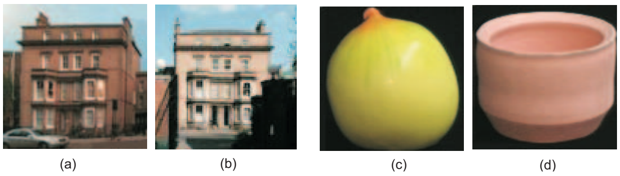

Fig. 1. Examples show that color is an important cue. The first two images cannot be discriminated if color information is not considered. No reliable interest points are detected in (c) and (d).

retrieval are parsed into words based on a vocabulary; then different weights are assigned to each term according to the frequency of the term in the document. Using the *k*-means algorithm [2], a visual vocabulary is constructed to realize the above ideas. In the visual vocabulary, each term is the centroid of a descriptor cluster. An inverted index is built based on this visual vocabulary [15].

The number of the interest points detected in an image depends on the size and the structure of the image. The image representation based on the interest points is not discriminative enough when few interest point is detected. Given the number of the interest points, different weights are assigned to the dissimilarity measures of interest points and edges at the image retrieval stage.

In section II, the integrated feature detection algorithm is proposed; indexing is discussed in section III; the dissimilarity measure methods are given in section IV; experiments based on two image databases are described in section V; the paper is concluded in section VI.

II. Feature Detection

The Harris detector [1] has been widely used in stereo matching, object detection and image retrieval for its repeatable detection performance. The basic idea of this detector is to use the autocorrelation function in order to determine locations where the signal changes in one or two directions. A matrix related to the auto-correlation function is computed:

$$C(\mathbf{x}, \sigma_I, \sigma_D) = \sigma_D^2 G(\mathbf{x}, \sigma_I) * \begin{pmatrix} L_x^2(\mathbf{x}, \sigma_D) & L_x L_y(\mathbf{x}, \sigma_D) \\ L_x L_y(\mathbf{x}, \sigma_D) & L_y^2(\mathbf{x}, \sigma_D) \end{pmatrix} \tag{1}$$

where σ_D is the derivation scale, σ_I the integration scale, G the Gaussian, and L the image smoothed by a Gaussian kernel.

This matrix has two eigenvalues that are the principal curvatures of the auto-correlation function. No structure exists when the two eigenvectors are very small. One large and one small eigenvalues indicate the presence of an edge. If both of them are very large and distinct, there is a corner-like structure. Edges and interest points can be computed based on:

$$det(C) - \alpha \cdot trace^2(C) < T_E, \tag{2}$$

and

$$det(C) - \alpha \cdot trace^2(C) > T_C. \tag{3}$$

Edges are computed based on on eq.(2), where α is the coefficient of the Harris function and T_E is the threshold of the Harris function($T_E < 0$). The edge detection is carried out at the first scale. Interest points can be detected by using eq.(3), T_C is the threshold for interest points($T_C > 0$).

The interest points detected by the Harris detector are not invariant to scale changes. The Harris-Laplace detector can detect scale invariant features [5]. The first step of this method is to compute interest points (Harris points) at different scales. Then the points with a local maximal measure (the Laplacian) are selected as Harris-Laplace interest points. According to [3], local extrema over scales of normalized derivatives indicates the presence of the characteristic local structures. The Laplacian is used in the Harris-Laplace interest point detection due to its high detection rate [5]. The scale of the point with the maximum of the Laplacian is taken as the characteristic scale of this interest point. The Laplacian function is defined as:

$$|\sigma^2(L_{xx}(\mathbf{x}, \sigma) + L_{yy}(\mathbf{x}, \sigma))|. \tag{4}$$

III. Indexing

Each image is represented by an edge color histogram and a VSM vector (description of interest points). The edge color histograms are indexed into a color information database and interest points are indexed into a VSM model based on the visual vocabulary learned from all the interest points.

A. Color Histogram Building and Indexing

The color histogram of an image is built by counting the number of pixels of each color. The most common RGB space is not used in this work because it is not perceptually uniform [13]. HSV is adopted here to characterize different colors because the transformation from RGB space is non-linear but easily invertible [13]. However, HSV brings about the disadvantage that hue is singular at the chromatic axis $r \approx g \approx b$ or $s \approx 0$. Three gray colors are introduced to overcome this problem. The color quantization algorithm in this work is similar to that of [13]. However, we use 12 hues, 2 saturation and 3 values. Including the three gray colors,the number of the colors is 75.

Two histograms are obtained using the statistic algorithm in [12]. In this work, The vector of the color histogram is normalized to unit length to reduce the scale effects. Since the dominant colors in an image play more important role than other colors which occupy relatively a small percentage of the image, the histograms are further refined by assigning zero to the values which are less than 15% of the maximum value. After the normalization, the color histogram reflects the distribution of colors instead of the absolute numbers of color pixels. The value of 15% is determined experimentally by testing on images containing differing colors.

B. Indexing of SIFT descriptors

The images in the database are indexed into the inverted index. This is an off-line process. Harris-Laplace interest points detected in these images are described by the SIFT descriptor that is an 128-dimension vector. The visual vocabulary is learned from these features by using the *k*-means algorithm. Based on this visual vocabulary, these descriptors are weighted and indexed into the inverted index using the VSM. Each image in the database is represented by:

$$\mathbf{d}_j = (w_{1,j}, w_{2,j}, \cdots, w_{n_t,j}), \tag{5}$$

where *w* is computed using the *tf-idf* scheme [15]. The details of the indexing algorithm can be found in [15].

IV. Retrieval

The retrieval of an input image is composed of two dissimilarity measures: color histogram dissimilarity measure and VSM vector dissimilarity measure.

A. Dissimilarity measure of color histograms

The measure methods for the dissimilarities between two histograms ($B = \{b_i\}$, histogram of an image in database; and $Q = \{q_i\}$, histogram of the query image) can be divided into two categories: bin-by-bin dissimilarity measures and cross-bin dissimilarity measures. The bin-by-bin measures only compare contents of corresponding histogram bins. The cross-bin measures also contain terms that compare non-corresponding bins [8]. The bin-by-bin measures are more sensitive to the position of bin boundaries. However, the cross-bin measures are computationally expensive. The histogram intersection which is a bin-by-bin measure is adopted in this work because of its ability to handle partial matches when the area of the two histograms are different [8]. Intersection values are normalized by the number of pixels in the model histogram, and thus matches are between 0 and 1. Histogram intersection is computed by:

$$d_o = 1 - \frac{\sum \min(b_i, q_i)}{\sum q_i}, \tag{6}$$

The dissimilarity of the two histograms is computed based on histogram intersections of the two directions (d_H and d_V):

$$s_{h,i}(B, Q) = \frac{w_H d_H + w_V d_V}{w_H + w_V}, \tag{7}$$

where w_H and w_V are weights assigned to the two intersections.

B. Dissimilarity measure of VSM vectors

The VSM evaluates the dissimilarity between an image in the database and the query image as the correlation between the two vectors $\mathbf{d}_j$ and $\mathbf{q}$.

$$\mathbf{q} = (w_{1,q}, w_{2,q}, \cdots, w_{t,q}), \tag{8}$$

The VSM assumes that the dissimilarity value is an indication of the relevance of an image in the database to the given query. Thus the VSM ranks the retrieved images by the dissimilarity value. In this work, the cosine of the angle between the two vectors is employed to measure the similarity between the query image and the image *j* in the database:

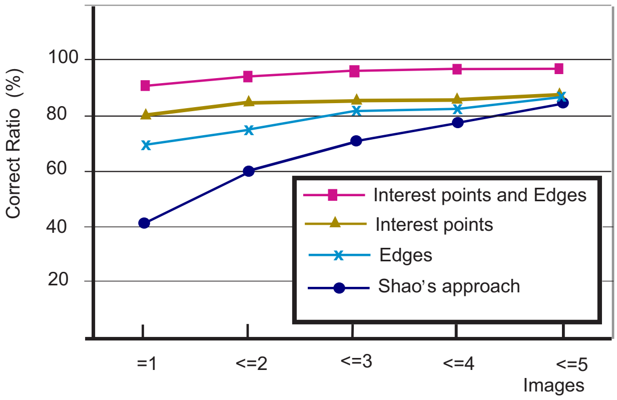

Fig. 2. Comparison of the correct ratio of our system with that of Shao [10]. Experiments are conducted on the ZuBuD image database. *y* (Vertical axis) is the correct ratio. The correct retrieval result is ranked within the top *x* (Horizontal axis) of the retrieved views.

$$s_{i,j} = 1 - \frac{\mathbf{d}_j^T \mathbf{q}}{\|\mathbf{d}_j\| \|\mathbf{q}\|}. \tag{9}$$

C. Combining the dissimilarity measures

The dissimilarities of eq.(7) and eq.(9) are incorporated to evaluate the overall dissimilarity between the input images and the image in the database.

$$s_j = \alpha s_{i,j} + \beta s_{h,i}, \tag{10}$$

$$\alpha + \beta = 1. \tag{11}$$

where α and β are coefficients of the dissimilarities. If many interest points are detected in the two images, eq.(9) plays a more important role than eq.(6), otherwise eq.(6) contributes more to the retrieval result. α:β is determined experimentally according to the number of the interest points(n_i) detected on the image and the size of the image.

V. Experiments

The image retrieval system based on edge color histograms and Harris-Laplace interest points has been tested on two image databases. First, we carry out tests on the COIL-100 object database [7]. COIL-100 is a popular image database for benchmark. Many objects in this database are so simple that only less than ten interest points are detected in one image. For each object, 12 views(view 0, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 330) are included in our database and another 60 views are used as query images. The recognition rate is 91.5%. This result is better than that of [11], in which the correct ratio is 59.5%.

The ZuBuD is a database containing 1005 images of 201 buildings in Zürich[9]. The size of the images is 640 × 480. Each building has been photographed from five different viewpoints. These images have been captured in different seasons and under different weather conditions. The query set consists of 115 images, which are not included in the

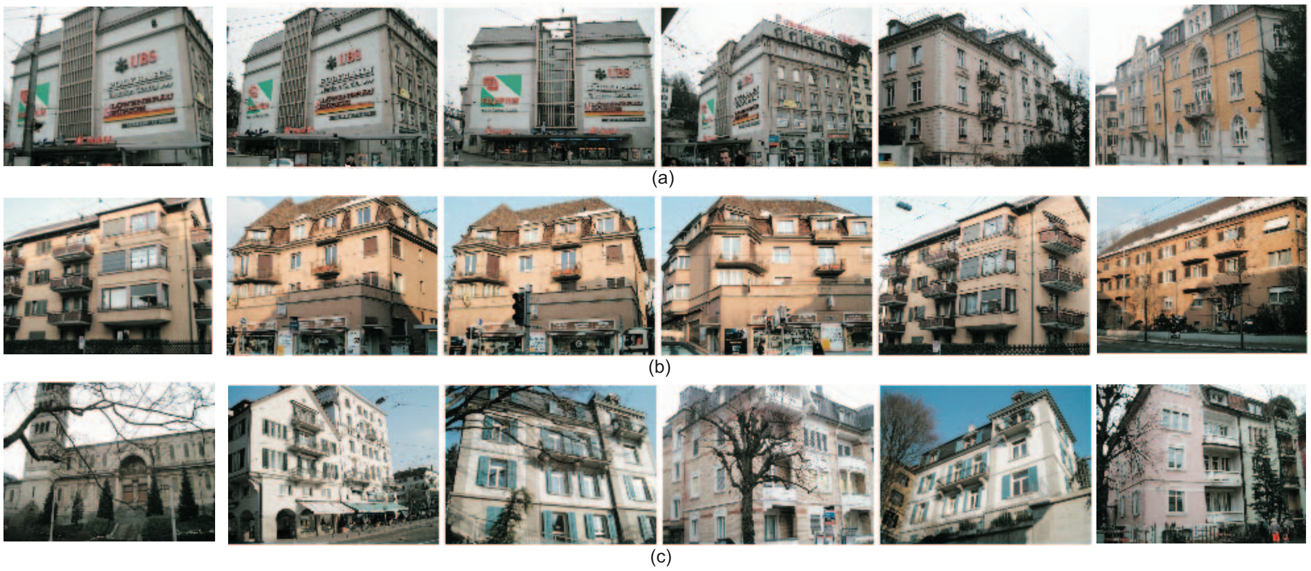

Fig. 3. Retrieval results from the ZuBuD image database. In each row, the first image is the query view, others are retrieval results with descending order of similarities. The retrieval results are correct in the first row and the second row. The retrieval result in the third row is not correct because of the serious occlusion of the tree.

database. Our method returns the correct match for 105 images. For an additional 5 images, it gives the correct match within the top five. The rest 5 images have not been recognized among the first five and are considered as failures. our result is much better than the result in [10]. Their result is improved in [11] by introducing a new indexing method - HPAT (Hyper-Polyhedron with Adaptive Threshold). Their retrieval system returns the correct match for 99 images. The correct matches are within the top five for additional 10 images. However, The recognition rate of our system (91.3%) is still better than theirs' (86%). In Fig.3(a), the result is found correctly at the first place. The correct result is ranked at the fourth position in Fig.3(b). Fig.3(c) shows one failure example in which the query image is occluded by a tree which brings about many outliers.

The edge color histogram index and the inverted index are built off-line. It takes 15 minutes to construct the two indexes of the 1005 images in the ZuBuD on an 1.4GHz laptop. For image retrieval, the computation time for the feature detection(including the detection of interest points and edges) is 0.55 ± 0.13 seconds for one query image on the laptop. It takes 0.19 ± 0.04 seconds to get the results from the database. Our approach is more efficient than that of [10].

VI. Conclusion

This paper discusses a novel content-based image retrieval approach based on edge color histogram and SIFT descriptors. The experimental results on the two databases demonstrate that this approach is efficient and reliable. The proposed method is also robust to viewpoint changes, partial occlusions, and partial illumination changes.

References

[1] C. Harris and M.J. Stephens. "A combined corner and edge detector". In *Proc. Alvey Vision Conference*, pp. 147-152, 1988.

[2] T. Kanungo, D.M. Mount, N. Netanyahu, C.D. Piatko, R. Silverman, and A. Y. Wu. "An efficient k-means clustering algorithm: analysis and implementation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(7), pp. 881-892, 2002.

[3] T. Lindeberg, "Feature detection with automatic scale selection", *Int'l. Journal of Computer Vision*, 30(2): pp. 79-116, 1998.

[4] D. G. Lowe, "Object recognition from local scale-invariant features", in *Proc. of Int'l Conf. on Computer Vision*, pp.1150-1157, 1999.

[5] K. Mikoljczyk and C. Schmid. "Indexing based on scale-invariant features", in *Proc. of Int'l Conf. on Computer Vision*, pp. 525-531. 2001.

[6] K. Mikoljczyk and C. Schmid. "A performance evaluation of local descriptors", in *Proc. of IEEE Int'l. Conf. on Computer Vision and Pattern Recognition*, pp. 1403-1410, 2003.

[7] S.A. Nene, S.K. Nayar and H. Murase. "Columbia Object Image Library (COIL-100)", TR CUCS-006-96, Dept. Comp. Sc., Columbia University, 1996.

[8] Y. Rubner, C. Tomasi, and L. J. Guibas. "The Earth Mover's Distance as a metric for image retrieval", *Int'l Journal of Computer Vision*, 40(2), pp. 99-121, 2000.

[9] H. Shao, T. Svoboda, and L.V. Gool. "ZuBuD – Zurich buildings database for image based recognition", *Technical Report 260*, Computer Vision Laboratory, Swiss Federal Institute of Technology, March 2003. Database downloadable from http://www.vision.ee.ethz.ch/showroom/.

[10] H. Shao, V. Ferrari, T. Svoboda, and L.V. Gool, "Fast indexing for image retrieval based on local appearance with re-ranking", in *Proc. of IEEE Int'l Conf. on Image Processing*, 2003.

[11] H. Shao, T. Svoboda, T. Tuytelaars, and L. Van Gool. "HPAT indexing for fast object/scene recognition based on local appearance", In *Int'l Conf. on Image Video Retrieval*, pp. 71-80, 2003.

[12] S-O. Shim, T-S Choi. "Edge color histogram for image retrieval", in *Proc. of IEEE Int'l Conf. on Image Processing*, pp. III957-960, 2002.

[13] J.R. Smith and S. Chang. "Single color extraction and image query", in *Proc. of IEEE Int'l Conf. on Image Processing*, pp. 528-531, 1995.

[14] M. Swain and D. Ballard. "Color indexing", *Int'l Journal of Computer Vision*, 7(1):11-32, 1991.

[15] J. Wang, R. Cipolla, and H. Zha. "Vision-based global localization using a visual vocabulary", to appear in *Proc. of IEEE Int'l Conf. on Robotics and Automation*, 2005.