World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:5, No:11, 2011

Local Image Descriptor using VQ-SIFT for Image Retrieval

Qiu Chen, Feifei Lee, Koji Kotani, and Tadahiro Ohmi

Open Science Index, Computer and Information Engineering Vol:5, No:11, 2011 waset.org/Publication/2562

Abstract**—**In this paper, we present local image descriptor using VQ-SIFT for more effective and efficient image retrieval. Instead of SIFT's weighted orientation histograms, we apply vector quantization (VQ) histogram as an alternate representation for SIFT features. Experimental results show that SIFT features using VQ-based local descriptors can achieve better image retrieval accuracy than the conventional algorithm while the computational cost is significantly reduced.

Keywords**—**SIFT feature, Vector quantization histogram, Local descriptor, Image retrieval.

I. Introduction

In recent years, local image descriptors are getting hot topics within the computer vision research area. Many local descriptors have been proposed [1]-[6] to achieve efficient object recognition and Content Based Image Retrieval (CBIR). The merits of these local descriptors are effective to partial occlusion, and are partially invariant to illuminations, projective transforms, and common object variations, etc., which make them suitable for employing in a lot of actual object recognition applications [7],[8] and CBIR systems [16].

There are roughly two main stages of processing while applying a local descriptor to object recognition application or CBIR system. The first is how to detect feature points (key-points) in the image. This is mainly the processing of the localization of key points and the determination of scale. The second is how to describe the key points which are invariant to rotation, projective transforms, and illuminations, etc. Although the two stages, detection and description of key points are necessary for the application of local descriptor, and are always proposed together by almost all local descriptors, they are independent problems actually. In another word, we can achieve the processing of two stages by different approaches.

In reference [9], the performances of several typical local descriptors [1]-[6], such as steerable filters [3], differential invariants [4], moment invariants [6] and SIFT [2], etc. have been investigated. It was found the accuracies of algorithms were relatively insensitive in key-point detection stage, but that were much different in description stage. SIFT algorithm which proposed by Lowe [2] obtained the best performance in matching experiments. In this paper, we pay attention to the second description stage, and propose a more robust local descriptor based on vector quantization (VQ) histogram [10].

Previously, Kotani et al. [10] have proposed a very simple yet highly reliable VQ-based face recognition method called *VQ histogram method* by using a systematically organized Codebook for 4x4 blocks with 33 codevectors having monotonic intensity variation without DC component.

VQ algorithm [11] is well known in the field of image coding (compression) and schematically. Input image is first divided into small blocks, which are taken as input vectors in VQ operation. Each input vector is then matched with codevectors in a codebook by calculating distances between them. The codevector having the maximum similarity to the input vector is selected by searching the minimum distance and the index number of the selected codevector is output.

This index number information is used for represent facial feature. It was found that a codevector histogram, which is obtained by counting the matching frequency of individual codevector, contains very effective facial feature information. By utilizing this technique, a novel face recognition algorithm called VQ histogram method has been developed.

The essence of VQ histogram method can be considered that the operation detects and quantizes the direction and the amount of intensity variation in the image block of the face. Hence VQ histogram also can effectively represent image feature information.

In this paper, instead of using SIFT's smoothed weighted orientation histograms, we apply vector quantization (VQ) histogram as an alternate representation for local image descriptor, which is more robust to rotation, projective transforms, and illuminations, etc. than the standard SIFT representation. We also evaluate proposed VQ-SIFT local descriptor in an image retrieval application using images taken from different viewpoints [15] with occlusions, and various lighting conditions.

The structure of this paper is as follows. In section II, the SIFT algorithm and Vector Quantization (VQ) histogram method will be introduced, and then proposed VQ-based representation will be described in detail in section III. Experimental results of image retrieval compared with the standard SIFT algorithm will be discussed in section IV. Finally, conclusions will be given in section V.

II. Related Works

A. Standard SIFT Algorithm

The origianl SIFT algorithm [2] consists of two main stages of processing. The first stage is detection of feature points (key points) in the image, which contains scale-space extrema detection and keypoint localization. The second is description of key points, which contains orientation assignment and keypoint descriptor.

Qiu Chen, Feifei Lee, and Tadahiro Ohmi are with New Industry Creation Hatchery Center, Tohoku University, Sendai, 980-8579 Japan (e-mail: qiu@fff.niche.tohoku.ac.jp).

Koji Kotani is with Department of Electronics, Graduate School of Engineering, Tohoku University, Sendai, 980-8579 Japan.

Figure 1 illustrates the calculation of the keypoint descriptor. First, as shown in figure 1(a), the coordinates of the descriptor and the gradient orientations are rotated relative to the keypoint orientation, thus orientation invariance is achieved. Then the image gradient magnitudes and orientations are sampled around the keypoint location, using the scale of the keypoint to select the level of Gaussian blur for the image. These are illustrated with small arrows at each sample location in figure 1(b).

A Gaussian weighting function with σ equal to one half the width of the descriptor window is used to assign a weight to the magnitude of each sample point. This is illustrated with a circular window in figure 1(b). These samples are then accumulated into orientation histograms summarizing the contents over 4×4 subregions, as shown in figure 1(c), with the length of each arrow corresponding to the sum of the gradient magnitudes near that direction within the region. Because there is 4×4 array of histograms with 8 orientation bins in each, a 4 × 4 × 8 = 128 element feature vector for each keypoint is utilized.

B. *Vector Quantization (VQ) Histogram*

Vector Quantization (VQ) histogram method [10] has been applied in face recognition application. Figure 2 shows process steps of VQ histogram method. First, low-pass filtering is carried out using simple 2-D moving average filter. This low-pass filtering is essential for reducing high-frequency noise and extracting most effective low frequency component for recognition. Block segmentation step, in which input image is divided into small image blocks (for example, 2×2) with overlap, namely, by sliding dividing-partition one pixel by one pixel, is the following. Next, minimum intensity in the individual block is searched, and found minimum intensity is subtracted from each pixel in the block. Only the intensity variation in the block is extracted by this process. This is very effective for minimizing the effect of overall brightness variations. Vector quantization is then applied to intensity-variation blocks (vectors) by using a codebook which prepared in advance. The most similar (matched) codevector to the input block is selected.

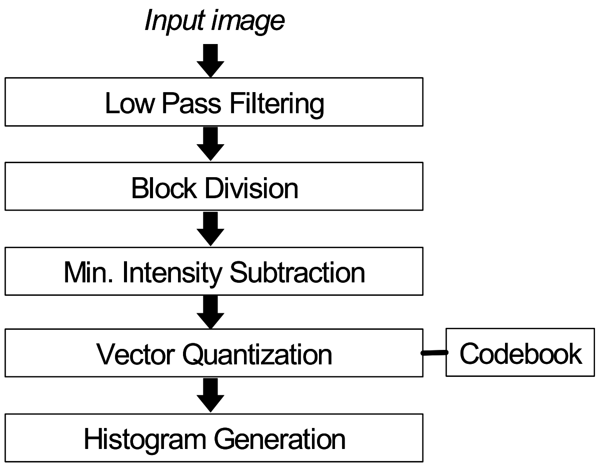
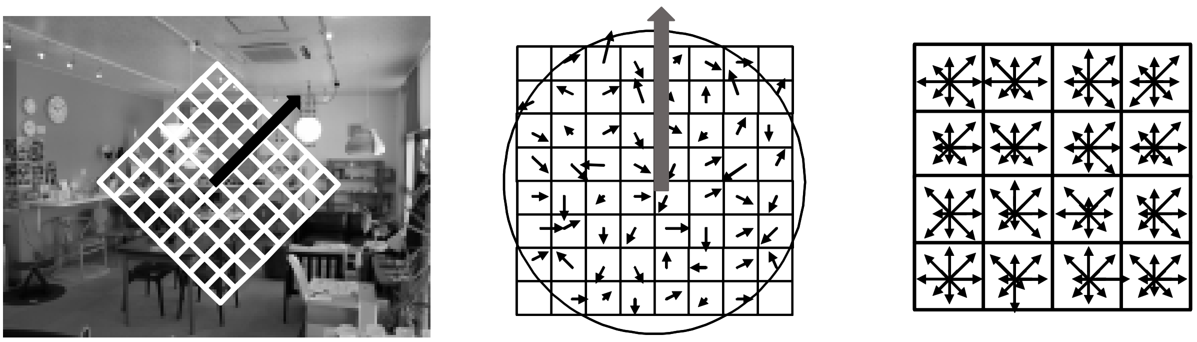


Fig. 2 Process steps of VQ histogram method

After performing VQ for all blocks divided from an input image, matched frequencies for each codevector are counted and histogram is generated. In the face recognition application, this histogram becomes the feature vector of the human face. Experimental results show recognition rate of 95.6 % for 40 persons' 400 images of publicly available face database of AT&T Laboratories Cambridge [14] containing variations in lighting, posing, and expressions.

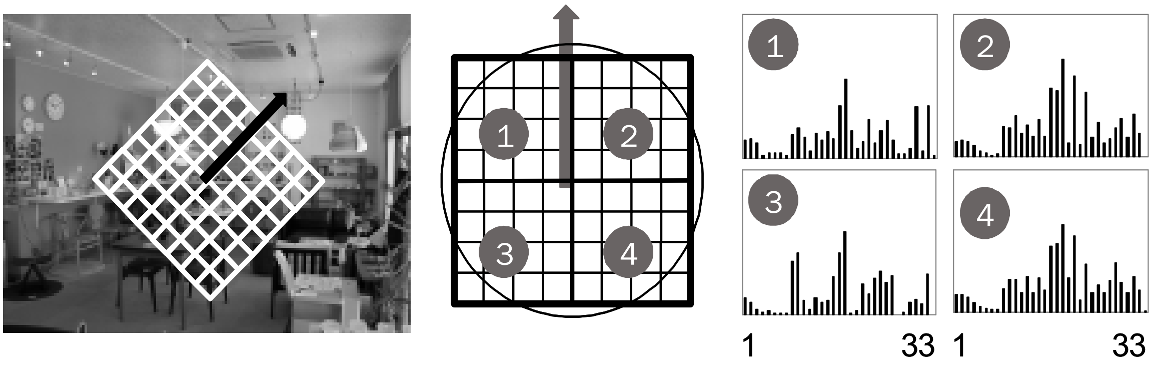
III. Local Descriptor Using VQ

For SIFT local descriptor, orientation histogram is utilized as the feature vector to represent the feature of keypoint. Previously, Freeman et al. [12] applied orientation histogram to gesture recognition. Experimental results showed that almost 10 types of gesture can be identified by using orientation histogram as feature vector. Other investigation [13] supported this result, and concluded that was the limit by using orientation histogram. That is to say, orientation histogram can only distinguish less than 10 types of variation. Although VQ histogram method also utilizes histogram information, the essence of the method is that the operation detects and



(a) (b) (c)

Fig. 1 Description of SIFT feature



(a) (b) (c)

Fig. 3 VQ-based local descriptor for SIFT. The coordinates of the descriptor and the gradient orientations are rotated relative to the keypoint orientation to keep the orientation invariance as shown in (a). Then the VQ histograms of 2x2 subregions are generated, as shown in (b), (c)

quantizes the direction and the amount of intensity variation in the image block. Hence VQ histogram can effectively represent image feature information and can be expected to extract more robust feature for keypoint.

Instead of SIFT's weighted orientation histograms, we apply vector quantization (VQ) histogram as an alternate representation for local image descriptor.

Our proposed algorithm for local descriptor utilizes the same input as standard SIFT descriptor, such as the sub-pixel location, scale, and gradient orientations of the keypoint, etc.

Figure 3 illustrates the calculation of the keypoint descriptor. First, as shown in figure 3(a), the coordinates of the descriptor and the gradient orientations are rotated relative to the keypoint orientation to keep the orientation invariance. Then the VQ histograms of 2x2 subregions are generated, as shown in figure 3(b), (c). Because the bin of each histogram is 33, 132 element feature vector for each keypoint is obtained.

IV. Experiments and Discussions

A. Image Data set

To compare the performance of our VQ-based local descriptor with the standard SIFT local descriptor, we utilize a dataset of 340 images containing various scenes, for example, objects, office, landscape, etc., which collected from internet. The size of each image is scaled to 640x480. We applied the following transformations to each image: (1) Gaussian noise (σ=0.05); (2) rotation of 45 degree followed by a 50% scaling; (3) intensity scaling of 50%.

B. Evaluation Method

The evaluation experiment of our VQ-based local descriptor and the standard SIFT local descriptor is implemented as follows.

(1) Firstly, all keypoints are extracted from both the original images and transformed images in the dataset.

(2) All pairs of keypoints from original images and transformed images are examined by using Euclidean distance as formula (1).

$$e(D_i)=\sqrt{\sum_{j=1}^{33} D_{i,j}^2}=\sqrt{\sum_{j=1}^{33}(I_j-R_{i,j})^2} \qquad i=1,2,\ldots,N_I \tag{1}$$

(3) If the Euclidean distance between the feature vectors for a particular pair of keypoints becomes smaller than the chosen threshold, this pair is termed a *match* (including *correct-match* and *false-match*). A *correct-match* is a match where the two keypoints correspond to the same physical location. A *false-match* is a match where the two keypoints come from different physical locations.

(4) Change the threshold, and determine *recall* and *1-precision* curve defined as follows.

$$\text{Recall}=\frac{\text{number of correct - match}}{\text{total number of positives}} \tag{2}$$

$$\text{1 - precision}=\frac{\text{number of false - match}}{\text{total number of matches}} \tag{3}$$

C. Comparison of Matching Processing

Fig. 4 shows the comparison results of matching experiments, where images were transformed in several variations. Figure 4(a) shows the results in the case of adding Gaussian noise. We can see that VQ-SIFT is much better at handling noisy images for almost all values of 1-precision. Figure 4(b) shows the results in the case of the geometric transforms, where target images were rotated by 45 degree and scaled by 50%. While both of the representations are not particularly well-suited to this task, VQ-SIFT appears more robust than the standard SIFT algorithm. Figure 4(c) shows that all of the representations are well-suited to variations in illumination.

From the comparison results above, we can conclude that the VQ-based local descriptors are more robust to image deformations than the standard SIFT algorithm.

D. Evaluation of Image Retrieval

We also evaluated proposed VQ-SIFT local descriptor in an image retrieval application using images taken from different viewpoints [15] with occlusions, and various lighting

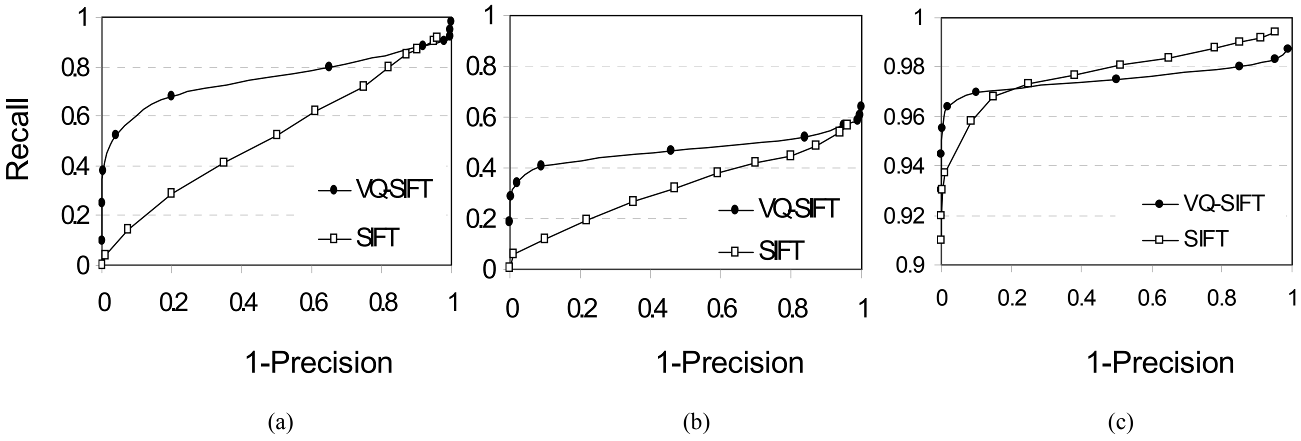


Fig. 4 Comparison results between SIFT and VQ-SIFT
(a) Add Gaussian noise (b) Rotation of 45 degree followed by a 50% scaling (c) Intensity scaling of 50%

conditions. There are 30 images including 10 common items in this dataset. Each image was used as a query into the database.

Image retrieval is implemented as follows. Firstly, we extract all corresponding feature vectors for given two images, and then we compare each feature vector in one image with all feature vectors in the other image and count the number of features that are smaller than a threshold. The similarity is calculated by the number of matches. If both of the other two images of the corresponding object were returned in the top 3 positions, the algorithm counts 2 points; and if only one of the correct matches were in the top 3 positions, it will counts only 1 point; else, it counts 0 point. At last, the counts will be divided by 60.

Compared with the rate of correct retrieval of about 43% obtained by original SIFT algorithm, 72% is achieved by using proposed VQ-SIFT algorithm. We can see that matching accuracy at the keypoint level also leads to better image retrieval results.

We also proposed a Content Based Image Retrieval (CBIR) system combined with HSV color features, Tamura texture descriptor[18], and VQ-SIFT features to reduce the complexity and matching time of SIFT features. First, HSV color features and the texture descriptor are used to reduce the amount of image candidates, and then VQ-SIFT algorithm is carried out to find similar candidates. The time spending of the same search with the algorithm using combined features is about 3 times faster than VQ-SIFT algorithm using the data set described above.

V. Conclusions

In this paper, instead of using SIFT's smoothed weighted histograms, we apply vector quantization (VQ) histogram as an alternate representation for local image descriptor. Experimental results demonstrated that the VQ-based local descriptors are more robust to rotation, projective transforms, and illuminations, etc. , and more suitable for image retrieval than the standard SIFT algorithm.

Acknowledgment

This research was partially supported by research grant from Support Center for Advanced Telecommunications Technology Research, Foundation (SCAT).

References

[1] C. Harris and M. Stephens, “A combined corner and edge detector,” *In Alvey Vision Conference*, 1988, pp. 147-151.

[2] D. G. Lowe, “Object recognition from local scale- invariant features,” *In Proceedings of International Conference on Computer Vision*, 1999, pp.1150-1157.

[3] W. T. Freeman and E. H. Adelson, “The design and use of steerable filters,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.13, no.9, 1991, pp. 891-906.

[4] J. Koenderink and A. van Doorn, “Representation of local geometry in the visual system,” *In Biological Cybernetics*, vol.55, 1987, pp. 367-375.

[5] F. Schaffalitzky and A. Zisserman, “Multi-view matching for unordered image sets,” *In Proceedings of European Conference on Computer Vision*, vol.1, 2002, pp. 414-431.

[6] L. Van Gool, T. Moons, and D. Ungureanu, “Affine/photometric invariants for planar intensity patterns,” *In Proceedings of European Conference on Computer Vision*, 1996.

[7] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol.60, no.2, 2004, pp. 91-110.

[8] R. Fergus, P. Perona, and A. Zisserman, “Object class recognition by unsupervised scale-invariant learning,” *In Proceedings of Computer Vision and Pattern Recognition*, Jun. 2003.

[9] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” *In Proceedings of Computer Vision and Pattern Recognition*, Jun. 2003.

[10] K. Kotani, Q. Chen, and T. Ohmi, “Face recognition using vector quantization histogram method,” *In Proceedings of 2002 International Conference on Image Processing*, vol. II of III:II-105-II-108, 2002.

[11] A.Gersho and R.M.Gray, *Vector Quantization and Signal Compression*, Kluwer Academic, 1992.

[12] W. T. Freeman and M. Roth, “Orientation histograms for hand gesture recognition,” *In Proceedings of International Workshop on Automatic Face and Gesture Recognition*, IEEE Computer Society, Zurich, Switzerland, 1995, pp. 296-301.

[13] http://phd.serkangenc.com/orientation/orientation.php.

[14] AT&T Laboratories Cambridge, The Database of Faces, at http://www.cl. cam.ac.uk/research/dtg/attarchive/facedatabase.html.

[15] http://www.cs.cmu.edu/~yke/pcasift/.

[16] T. Deselaers, D. Keysers, and H. Ney, “Features for image retrieval - a quantitative comparison”, 26th DAGM Symposium, vol. 3175 of Lecture Notes in ComputerScience , Germany, pp. 228-236, 2004.

[17] Q. Chen, K. Kotani, F. F. Lee, and T. Ohmi, “Robust VQ-based Local Descriptor for SIFT Feature,” *Proceeding of the International Conference on Image and Vision Computing (ICIVC 2009)*, pp. 1329-1333, France, Jun. 2009.

[18] H. Tamura, S. Mori, and T. Yamawaki, “Textural features corresponding to visual perception”, IEEE Transaction on Systems, Man, and Cybernetics, vol. 8, no. 6, 460–472, 1978.