An Improved Feature Vector for Content-based Image Retrieval in DCT Domain

Cong Bai, Kidiyo Kpalma and Joseph Ronsin

Université Europénne de Bretagne, INSA de Rennes, IETR, UMR 6164, Rennes, France

Keywords: Content-based Image Retrieval, DCT, Texture, Face Recognition.

Abstract: This paper proposes an improved approach for content-based image retrieval in Discrete Cosine Transform domain. For each 4x4 DCT block, we calculate the statistical information of three groups of AC coefficients and propose to use these values to form the AC-Pattern and use DC coefficients of neighboring blocks to construct DC-Pattern. The histograms of these two patterns are constructed and their selections are concatenated as feature descriptor. Similarity between the feature descriptors is measured by χ^2 distance. Experiments executed on widely used face and texture databases show that better performance can be observed with the proposal compared with other classical method and state-of-the-art approaches.

1 INTRODUCTION

As a transform adopted in JPEG compression standard, Discrete Cosine Transform (DCT) is a powerful tool to extract features from images. Consequently in last decades, many researches appeared in image retrieval based on DCT. DCT has the capability to compact the energy, i.e. much of the energy lies in low frequency coefficients, so high frequency can be discarded without visible distortion. In other words, a reduced part of DCT coefficients can efficiently represent the image contents. In comparison to the use of all of the coefficients, this consideration reduces the complexity and redundancy of the feature vectors that are generated for image retrieval. Furthermore, the AC coefficients of some regions represent some directional information.

Different compositions of the coefficients for a feature vector construction have been proposed. In (Tsai et al., 2006), the upper left coefficients of each 8x8 block are categorized into four groups: one is DC coefficient and other three includes the AC coefficients which have vertical, horizontal and diagonal information. These four groups compose the feature vectors. In (Zhong and Defée, 2005), 16 coefficients from each 4x4 block are used to construct 2 patterns and then their histograms are constructed as descriptors to do face recognition. In (Bai et al., 2012), a selection of coefficients is done and the feature descriptors are constructed from the histogram of this selection. Our approach proposed in this paper is in-spired from these previous works and those considerations mentioned in last paragraph.

In this paper, we present a simple but effective way to construct the feature vectors, and use the part of histogram of these vectors as the descriptors of the images. Similarity of the descriptors between query and images in the database is measured by χ^2 distance. Experimental results show that the proposed method can apply both on face database and texture database and can achieve better performance than the referred methods including state-of-the-art approaches.

The rest of the paper is organized as follows: the principle of constructing feature vectors is presented in section 2. Section 3 gives the process of forming feature descriptors and selection of similarity measurement. Experimental results are shown in section 4 and conclusion is given in section 5.

2 FEATURE VECTOR

Images are transformed by DCT firstly. In this study, we use 50% overlapping 4x4 DCT block transform. To eliminate the effect of luminance variations, luminance normalization needs to be done before constructing AC and DC patterns. We adopt the method presented in (Zhong and Defée, 2005) as a pre-processing step. From these pre-processed coefficients, AC-Patterns and DC-Patterns are constructed.

For each block, AC-Pattern is formed by 9 coefficients and DC-Pattern is constructed by the DC coefficients of the block itself and those of its 8 neighboring blocks. So intra-block features and inter-block features can be represented by these two patterns.

2.1 AC-Pattern

In DCT domain, energy is packed into a few coefficients, and some coefficients represent directional information. So the proposed approach selects 9 coefficients out of all 15 AC coefficients in each block and uses their statistical information to construct the AC-Pattern. These 9 coefficients are categorized into 3 groups: horizontal (Group H), vertical (Group V) and diagonal (Group D). For each group, the sum of the coefficients is calculated firstly and then the squared-differences between each coefficient and the sum of this group are calculated. Finally, the sums of squared-differences of each group are used to construct AC-Pattern. The process of forming AC-Pattern is shown in Figure 1. Compared with the method of (Zhong and Defée, 2005), this way of constructing AC-Pattern reduces obviously the complexities of the feature vector. For example, the maximal dimension of the AC-Pattern of (Zhong and Defée, 2005) is 15 and the one of ours is only 3. This number remains the same as the method of (Bai et al., 2012).
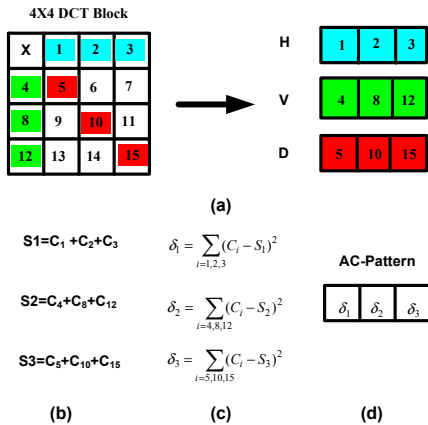

Figure 1: **AC-Pattern construction:** (a) Three groups of AC coefficients are extracted from DCT block (b) Sum of group (c) Sum of squared-differences (d) AC-Pattern.

2.2 DC-Pattern

Different from previous AC-Patterns that describe local information inside each block, DC-Pattern will consider global information by using gradients between DC coefficients of each block and those of its neighbors blocks. So DC-DirecVec (Zhong and Defée, 2005) is used as DC-Patterns.

3 FEATURE DESCRIPTOR AND SIMILARITY MEASUREMENT

3.1 Feature descriptor

In this study, we use the histogram of patterns instead of histogram of individual DCT coefficients as feature descriptor. The histogram of AC-Patterns and DC-Patterns are defined as the number of appearance of patterns in DCT domain. A disadvantage of the histogram method is that it requires a large number of histogram bins, typically several thousands, to capture accurately information of feature vector. Thus it leads to complexity in both storage of image indices and retrieval timing. To overcome this drawback, we adopt two modifications.

From the original histogram of AC-Patterns we can make two observations and then two improvements are respectively adopted: the first is that there is only a few part of AC-Patterns which appears in large quantities and a large number of AC-Patterns that appears rarely (Zhong and Defée, 2005). So in consideration of time-consuming and efficiency, we just select *ACbins* AC-Patterns which have highest occurrence to construct the histogram. The second observation is that the first AC-Pattern inside the histogram is very dominant. This AC-Pattern corresponds to uniform blocks in the image and consequently we will discard it. Thus we obtain the AC-Pattern histogram H_{AC}, as show in Figure 2. In this histogram, $ACbins = 70$.
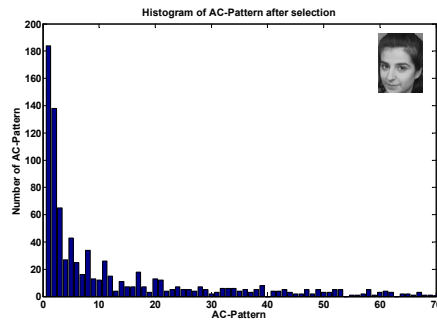

Figure 2: Histogram of AC-Patterns after selection.

For DC-Pattern histogram, we can also make the same observation that only a part of DC-Patterns occurs in large quantities while other part of DC-Patterns appears rarely. So we select *DCbins* dominant DC-Patterns to construct DC-Pattern histogram H_{DC}.

Finally, we use the concatenation of AC-Pattern and DC-Pattern histograms to build the descriptors. In this context, the descriptor is defined as follows:

$$D = [(1-\alpha) \times H_{AC}, \alpha \times H_{DC}] \quad (1)$$

where α is a weight parameter that controls the impact of AC-Patterns and DC-Patterns histogram.

3.2 Similarity Measurement

The similarity between query and images in the database is assessed by the distance between feature descriptors. A similarity measure assigns a lower distance or high score to more similar objects.

In our approach, χ^2 distance is used to measure the similarity. Assuming D_i and D_j are the query and target feature descriptors respectively, the χ^2 distance is defined as follows:

$$Dis_{i,j} = \sum_{k=1}^{m} \frac{(D_i(k) - D_j(k))^2}{D_i(k) + D_j(k)} \qquad (2)$$

where m indicates the dimension of descriptors.

4 EXPERIMENTAL RESULTS

To evaluate the proposal, we perform experiments on ORL face database (AT&T Laboratories Cambridge, 1992), GTF face database (Georgia Tech, 1999), FERET face database (Phillips et al., 2000) and Vis-Tex texture database (Media Laboratory, 1995).

4.1 Experiments on Face Databases

The ORL database includes 10 different images of 40 persons. The GTF database includes 15 different face images of 50 different persons. As we want to compare our proposal with the methods presented in (Zhong and Defée, 2005) and (Bai et al., 2012), we implemented similar experiments as them, that means use first 6 images as image database and remaining 4 images as query images for recognition on ORL database and first 11 images of each person as image database and remaining 4 images as query images on GTF database.

As ORL and GTF are relatively small databases, to further illustrate the contribution of the proposal, we also implement it on FERET database which contains more than 10000 images from more than 1000 persons. The two sets of frontal view faces **fa** and **fb** were selected to evaluate the proposed method: **fb** is used as query images for retrieval from the **fa**.

For evaluating the performance, we use Equal Error Rate (EER) (Bolle et al., 2000). Images are considered as similar if the distance between their features descriptors is under a given threshold. So considering a query image belonging to class A, two things could occur: on one hand, it could be recognized rightly; on the other hand, it could be falsely

rejected from class A, then the ratio of how many images of class A are in this situation is called False Rejected Rate (FRR). In contrast, considering a query image out of class A, when it is compared with the images of class A, it could be rejected rightly or it could be falsely accepted as class A, then the ratio of how many images of other classes are in this situation is defined as False Accept Rate (FAR). These two rates will change when the threshold changes. When FRR and FAR take equal values, an equal error rate (EER) is got. The lower the EER is, the better is the performance.

As mentioned before, the concatenation of the AC-Pattern histogram and the DC-Pattern histogram is used to do image retrieval. To verify the effectiveness of the proposed new method on constructing AC-Pattern, the DC-Patterns histogram is the same through all the experiments, but the construction of AC-Pattern are different. We name the method presented in (Zhong and Defée, 2005) as "Zhong", the method presented in (Bai et al., 2012) as "Bai". For these two methods and our proposal, we tested different sets of parameters to find the one that can assure the best performance, and only the best performances are compared that are shown in Table 1. From this table, we can see that the proposal outperforms the referred methods.

Table 1: Comparison of EER (%) on face databases.

Database/Method	Zhong	Bai	Proposal
ORL	6.07	3.75	**3.25**
GTF	11.57	11.19	**9.84**
FERET	3.75	3.57	**3.35**

4.2 Experiments on Texture Database

In order to evaluate extensibility of the proposal in a wider application field, we work on a selection of images from the Vistex Texture Database (Media Laboratory, 1995), consisting of 40 textures which have already been extensively used in texture image retrieval literature (Do and Vetterli, 2002) (Kokare M. and Chatterji, 2005) (Kwitt and Uhl, 2010).

The 512×512 pixels color version of the textures are divided into 16 non-overlapping subimages (128×128 pixels) and converted to gray scale images, thus creating a database of 640 images belonging to 40 texture-classes, each class includes 16 different samples. In this retrieval experiments, each image is used as query image. The relevant images for each query consists of all the subimages from the same original texture. Like in other literatures (Do and Vetterli, 2002) (Kokare M. and Chatterji, 2005) (Kwitt and Uhl, 2010), we use the average retrieval rate

(ARR) to evaluate the performance. For a given query image, and a given set of retrieved images, the retrieval rate is defined as the percentage of the number of correct images retrieved on the total number of retrieved images. For comparison purpose, top-16 match images are retrieved for each query.

Table 2: Comparison of ARR on Vistex.

Method	RCWF	CWT	PTR	CWT+RCWF	Proposal
ARR(%)	75.78	80.78	81.73	82.34	**84.04**

Table 2 provides an objective comparison of ARR. In this table, RCWF indicates Rotated Complex Wavelet Filters method proposed in (Kokare M. and Chatterji, 2005). CWT represents the Complex Wavelet Transform method presented in (Kingsbury, 1999). CWT+RCWF is a mixed method also presented in (Kokare M. and Chatterji, 2005). PTR (Kwitt and Uhl, 2010) is a probabilistic texture retrieval method based on dual-tree complex wavelet transform. From this table, it can be observed that our proposal outperforms other methods.

5 CONCLUSIONS

We have presented an improved approach for content-based image retrieval in DCT domain. A new way to construct AC-Patterns is proposed: this proposal uses the statistical information of every directional groups of AC coefficients. Compared with other methods of constructing AC-Patterns (Zhong and Defée, 2005), the proposed method reduces the dimension of AC-Patterns. And we also evaluated the extensibility of our proposal by applying it on two different kinds of database: face database, which has structural contents, and texture database, which has both structural and unstructured contents. Our approach is evaluated on widely used ORL, GTF, FERET and VisTex database. The experimental results show that our approach outperforms the referred methods, including state-of-the-art methods.

REFERENCES

AT&T Laboratories Cambridge (1992). ORL database. http:// www.cl.cam.ac.uk/research/dtg/attarchive/face database.html. Online; accessed March 2010.

Bai, C., Kpalma, K., and Ronsin, J. (2012). A new descripor based on 2D DCT for image retrieval. In *Proceedings of the International Conference on Computer Vision Theory and Applications*, pages 714–717.

Bolle, R., Pankanti, S., and Ratha, N. (2000). Evaluation techniques for biometrics-based authentication systems (FRR). In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, pages 831 –837 vol.2.

Do, M. and Vetterli, M. (2002). Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance. *Image Processing, IEEE Transactions on*, 11(2):146 –158.

Georgia Tech (1999). GTF database. http:// www.anefian.com/research/face_reco.htm. Online; accessed March 2010.

Kingsbury, N. G. (1999). Image processing with complex wavelet. *Phil. Trans. Roy. Soc.*, 357:2543–2560.

Kokare M., B. P. and Chatterji, B. (2005). Texture image retrieval using new rotated complex wavelet filters. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 35(6):1168–1178.

Kwitt, R. and Uhl, A. (2010). Lightweight probabilistic texture retrieval. *Image Processing, IEEE Transactions on*, 19(1):241 –253.

Media Laboratory, M. (1995). Vistex database of textures. http:// vismod.media.mit.edu/vismod/imagery/VisionTexture/. Online; accessed Dec. 2010.

Phillips, P., Moon, H., Rizvi, S., and Rauss, P. (2000). The FERET evaluation methodology for face-recognition algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(10):1090 – 1104.

Tsai, T., Huang, Y.-P., and Chiang, T.-W. (2006). Image retrieval based on dominant texture features. In *Industrial Electronics, 2006 IEEE International Symposium on*, volume 1, pages 441 –446.

Zhong, D. and Defée, I. (2005). DCT histogram optimization for image database retrieval. *Pattern Recognition Letters*, 26(14):2272 – 2281.