

# PATTERN RECOGNITION IN COMPRESSED DCT DOMAIN

*Daidi Zhong, Irek.Defée*

Institute of Signal Processing  
Tampere University of Technology  
P.O. Box 553,  
FIN-33101 Tampere , FINLAND.  
E-mail: daidi.zhong@tut.fi irek.defee@tut.fi

## ABSTRACT

Images and video are currently predominantly handled in compressed form. Block-based compression standards are by far the most widespread. It is thus important to devise information processing methods operating directly in compressed domain. In this paper we investigate this possibility on the example of simple face information processing method based on the DCT (Discrete Cosine Transform) blocks. We use patterns of quantized 4x4 DCT blocks for representing local picture information. These patterns at different quantization levels provide very flexible representation of picture information. We represent global information in pictures by histograms of quantized DCT pattern distributions. The approach is tested on database of face images and it is shown that despite its simplicity provides good results in the face recognition problem.

## 1. INTRODUCTION

Images and video are handled nowadays mostly in compressed format. This is due to the reduced size of the data which diminishes the need for storage and communication bandwidth. The compression methods used are highly optimized for the removal of all non relevant information while preserving high perceptual quality. Compression leaves only perceptual content and this is desirable not only from the data size reduction point but also for the visual information extraction. Hence visual information extraction in compressed domain should be advantageous but this point has not been widely followed due to the lack of proper approach. Such an approach is proposed in this paper and initial study illustrates certain important aspects of it. We show that

recognition results are very good considering the simplicity of method and that with adaptive selection of quantization results are excellent.

In this paper we study the problem of information extraction in compressed domain on the example area of face detection and recognition. This topic has been extensively studied in the past from a large variety of viewpoints. Face detection and recognition technology has also been widely used in many fields, such as personal identification, video surveillance system, human computer interaction, etc. In recent years, many peoples are working in this field to find efficient ways to recognize face images fast and accurately [1],[2]. Recently face information became important in multimedia information retrieval systems. Instead of the original linguistic description of images using some key words, these methods is more powerful for image indexing and face recognition, as it automatically considers the information from images themselves. This information is so-called visual features, a number of key visual descriptors of facial content, which can distinguish one face from other faces.

Generally, face information processing is based on extraction of local visual features, such as color, texture, shape and integrating them into data structure describing face. Recently simple histogram based methods were proposed for the face detection and recognition problems [3],[4]. In these methods histogram of basic features extracted is formed. Despite of simplicity of this approach results are surprisingly good. In [3] histogram method is applied to features extracted by block Vector Quantization (VQ). In [4] histograms based on skin color features were applied for face detection problems. Both these approaches give good results considering simplicity of methods used.

In this paper we use histogram method with feature extraction based on 4x4 DCT blocks. This has advantage

of enabling direct extraction in the compressed domain where 8x8 DCT is used. Scaling from the 8x8 DCT size to 4x4 DCT is straightforward and has been used in the past for picture downsizing [5]. Using the DCT for feature extraction has many advantages. The transform is easy to calculate, it is well preserving perceptual information under quantization in wide range of values. Using the DCT of size 4x4 is good compromise between the amounts of detail block size; same size is often used in application of vector quantization to image compression [4]. The difference between the VQ and DCT approach lays in much more flexible control of feature extraction by quantization and good properties of DCT in matching perceptually important features.

The approach presented illustrates the strength of the extraction of shape information in the DCT domain on the example of face recognition.

## 2. FEATURE DESCRIPTION USING DCT

For the feature extraction we use 4x4 DCT blocks. The 2-D DCT can be calculated directly by:

$$G(m, n) = a(m)a(n) \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} g(i, k) \cos\left[\frac{\pi(2i+1)m}{2N}\right] \cos\left[\frac{\pi(2k+1)n}{2N}\right] \quad (1)$$

$$a(0) = \sqrt{\frac{1}{N}} \quad \text{and} \quad a(m) = \sqrt{\frac{2}{N}}, \quad 1 \leq m \leq N$$

Here,  $g$  is the source block and the  $G$  is the DCT transformed block.  $N$  is the dimension of the blocks.

In compression standards 8x8 DCT is commonly used. The 4x4 DCT coefficients can be obtained from the 8x8 DCT blocks. One possible way is as shown as below:

$$DCT(B') = DCT(H)DCT(B)DCT(H^T) \quad (2)$$

Here,  $H^T$  is the transpose matrix of  $H$ .  $B$  is the 8x8 source block and the  $B'$  is the 4x4 output block.  $H$  is defined as below, [5]

$$H = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 \end{bmatrix} \quad (3)$$

In compression applications the DCT is quantized. Quantization eliminates very effectively non-perceptual information overhead. 8x8 blocks of DCT coefficients are quantized using quantization matrices. Methods for deriving quantization matrices based on the properties of human visual system were also proposed [7]. The 4x4 DCT quantization has been studied only recently in the context of new H.264 video compression standard [8]. In

this standard scalar quantization, so-called QP factor is applied to the coefficients of DCT blocks. For the purpose of this paper the QP factor is adopted as quantization method.

Elimination of perceptually irrelevant information by quantization can be considered desirable from the pattern recognition viewpoint. There is however a question what is the impact of QP factor in this regard. The effect of quantization can be seen as reducing the variety of DCT blocks, the number of different blocks images is reduced when the QP value increases. This point is illustrated in Fig. 1 where this relation is shown for typical picture.

It can be expected that for certain range of QP values recognition based on the DCT blocks will be facilitated since the number of blocks will be reduced while relevant information will be still preserved.

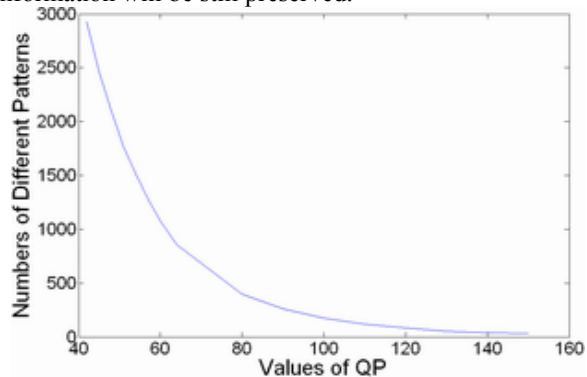


Fig.1 QP vs. Number of Patterns

## 3. APPLICATION TO PATTERN RECOGNITION

We will apply quantized DCT blocks to pattern recognition. The main idea is to study the effect of quantization on the recognition performance. As emphasized above increasing quantization results, up to a point, in the elimination of perceptually non-relevant information. Obviously, too strong quantization will remove relevant information too. Thus, there should be optimal level of quantization maximizing pattern recognition performance. Moreover, the overall performance should be good because of the feature-preserving properties of the DCT.

To check this reasoning we adopted simple pattern recognition scheme based on histograms [3]. For a database  $D$  of pictures  $D=\{P_1 \dots P_n\}$  histograms of quantized DCT blocks are calculated for each picture  $P_i$ . Pattern recognition for pictures in the database is based on comparison of histograms, using sum of absolute differences measure between the histograms. Minimum absolute difference (MAD) corresponds to best recognition candidate

$$MAD_i = \min |H_i - H_j| \quad j=1 \dots n \quad (4)$$

Fig. 2 shows the recognition process. Firstly, we divide the image into 4x4 blocks. Then we perform 4x4 DCT transform to them. If the source image has already been compressed into 4x4 DCT blocks, such as H.264, we can use these DCT blocks directly. The third step is quantization, aiming to limit the coefficients to a small range. After that, we can use some generic patterns to match the face images, count for each pattern to achieve a histogram, which representing the times of each pattern appearing in this image. At the registration period, the histogram of each input image is calculated at certain QP. At the searching period, we process the target image in the same way and calculate the difference of its histogram to other histograms. The image in the database, which has the lowest difference to the target one, is our final choice.

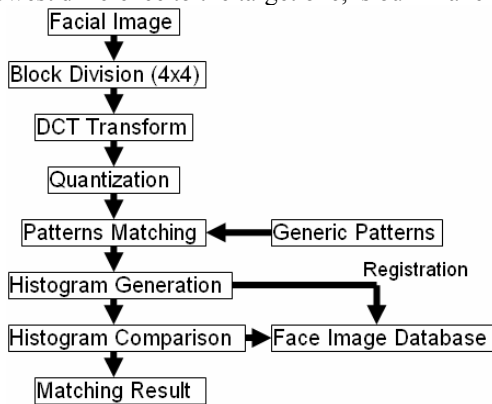


Fig.2 Face Recognition Process

### 5.1. 4x4 DCT blocks and quantization

In order to extract the local pattern information, we focus on the 15 AC coefficients of the 4x4 DCT block. After quantization, the 15 AC coefficients are limited to certain numbers of patterns and the total number of the patterns in a picture is dependent on the quantization factor QP. The type of source images has also impact on the type of DCT patterns. If the number of patterns is too small, then the difference of histogram between each image will also be too small for us to distinguish them; but if it is too big, the computation complexity is quite large. The overall

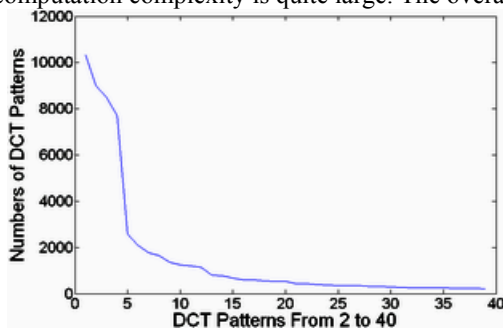


Fig.3 Probability of Patterns (QP=21)

question is about the dependence of performance on the number of patterns.

We can generate different patterns from a number of images, but a lot of them scarcely appear in other images. Fig.3 shows the frequency of these patterns appearing in typical image. For pattern recognition a set of patterns appearing most often can be selected. This set will depend on the type of picture, quantization factor QP, size of the picture database and required recognition performance.

The optimization problem is thus given a an image database D find quantization factor QP and DCT patterns set providing best recognition performance based on the histograms of DCT blocks and minimum absolute difference measure. One can notice that both best QP and DCT pattern set may be image- dependent. The question is what ultimately achievable performance of such system is. The performance has been studied in this paper on the example of face recognition from a face database.

## 4. EXPERIMENTAL SYSTEM AND RESULTS

A system for testing ultimate performance for face recognition problem based on histograms of quantized DCT blocks was developed. Block DCT transformation of images are calculated and quantized. Next a set of DCT patterns is selected for each image providing histogram for which the number of images found by the MAD measure is minimized as a function of the DCT pattern set and quantization factor QP.

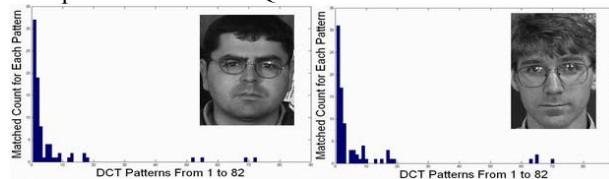


Fig.4 Histograms of Patterns For Two Persons

For experiments we used the Georgia Tech Face (GTF) Database [6]. We selected 3 images for each person, which are all front face images. Therefore, totally we got 150 images in our database. Fig.4 shows two

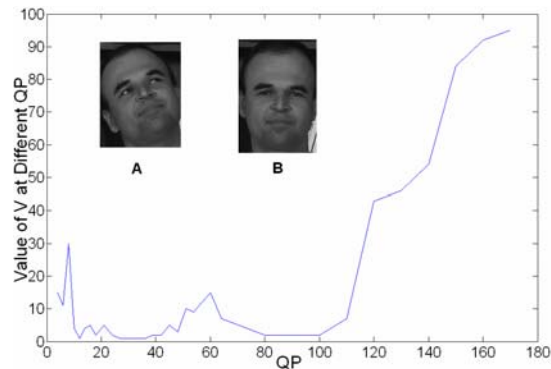


Fig.5 Searching for Fixed Number of Patterns

histogram of patterns for different persons.

If we use one target image to search among other 149 images in this database, we use a value  $V_i$ ,  $\{1 \leq V_i \leq 149, i=1, 2, \dots, 149\}$ , to evaluate the searching results.  $V_i = 1$  means that there is only one image which has the difference of histogram less or equal than the  $i$ th image. So,  $V_i = 1$  is the best result for  $i$ th image;  $V_i = 149$  is the worst result.

When the number of patterns is fixed to certain value, we can evaluate the value of  $V_i$  as a function of quantization factor QP. This is shown in Fig. 5 indicating that there are local flat minima in the function. QP can be thus selected in broad range of values for good results.

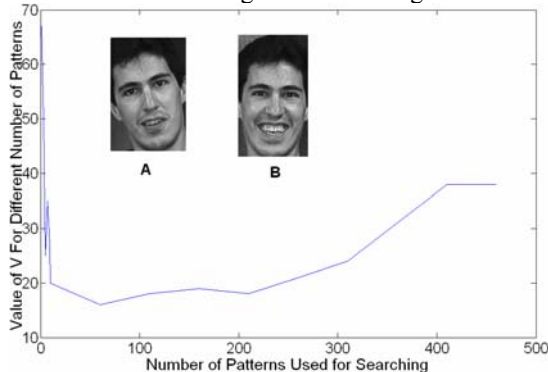


Fig.6 Searching for Fixed QP

In the second test, we fix the quantization factor QP and change the number of patterns. The result is shown in Fig. 6 indicating for smooth behavior with the existence of minimum.

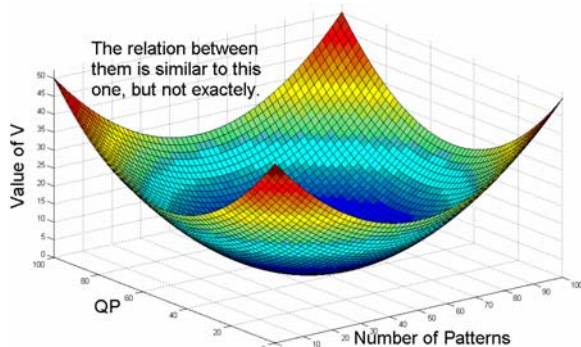


Fig.7 The Approximate Optimal Point

In the third experiment we perform the searching by varying both QP and pattern number. As shown in Fig. 7 we found that, the relation between result, QP and pattern number is as a whole a concave function. This indicates that there is global minimum resulting in very good database search result but depending on both QP and the number of patterns. For most face images, the best performance is obtained when QP is set to from 18 to 36.

For illustration, we select QP = 21, and use 30 basic patterns, with one target image A1 to search for images

A2 and A3, which are all taken from the same person in the database. The result shows that we can always find the right person. It means that either A2 or A3 or both has  $V_i \leq 12$  (8% of 150). The best match is dependent on the detail of this target image, such as face expression, hair type, beard type and background, etc. Table.1 shows us the searching result for two persons A and B:

Table 1. Searching Result for Person A and B

	A <sub>1</sub> -A <sub>2</sub>	A <sub>1</sub> -A <sub>3</sub>	B <sub>1</sub> -B <sub>2</sub>	B <sub>1</sub> -B <sub>3</sub>
$V_i$	9	1	1	2

\* A<sub>1</sub>-A<sub>2</sub> means: Use A<sub>1</sub> to search for A<sub>2</sub>.

## 5. CONCLUSIONS

In this paper, it is shown that quantized DCT and selection of proper pattern set results in very informative description for pattern recognition. The approach is illustrated using histograms of quantized 4x4 DCT blocks on face database images. By adjusting quantization level and DCT pattern set very good face recognition results are obtained. The method is computationally efficient and can be used directly for information extraction from compressed video. Further research is needed for complementing histograms with other measures of global pattern similarity based on histograms, and optimizing the patterns set by normalizing the bins of histograms.

## 11. REFERENCES

- [1] Ara V. Nefian, "Embedded Bayesian Networks for Face Recognition," in *IEEE International Conference on Multimedia and Expo*, vol. 2, p. 133-136, August 2002.
- [2] F. S. Samaria, "Face Recognition using Hidden Markov Models," PhD thesis, University of Cambridge, 1994.
- [3] Koji Kotani, Chen Qiu, and Tadahiro Ohmi, "Face Recognition Using Vector Quantization Histogram Method," in *International Conference on Image Processing*, II-105, Sep. 2002.
- [4] Dapang Chen, and Alan C. Bovik, "Visual Pattern Image Coding," in *IEEE Trans. on Communications*, pp. 2137, Dec. 1990.
- [5] Shih-Fu Chang; Messerschmitt, D.G., "Manipulation and compositing of MC-DCT compressed video", *IEEE Journal on Selected Areas in Communications*, Jan. 1995.
- [6] Georgia Tech Face Database, Available: <ftp://ftp.ee.gatech.edu/pub/users/hayes/facedb/>.
- [7] Joint Video Team(JVT) of ISO/IEC MPEG&ITU-T VCEG, "Draft ITU-T Recommendation and Final Draft International Standard Joint Video Specification", Document JVT-G050, March 2003
- [8] H. A. Peterson, H. Peng, J. H. Morgan, W. B. Pennebaker, "Quantization of color image components in the DCT domain", in *Human Vision, Visual Processing, and digital Display II, Proc. SPIE*, vol. 1453, pp. 210-222, 1991.