Proceedings of the 6th Nordic Signal Processing Symposium - NORSIG 2004
June 9 - 11, 2004
Espoo, Finland

Pattern Recognition by Grouping Areas in DCT Compressed Images

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 almost exclusively handled in compressed formats based on quantized block DCT transform. Information extraction from images and video has been traditionally studied in the pixel domain. At present methods operating in the DCT domain are more natural and required. There is also argument for DCT based information extraction based on efficiency: compressed images preserve perceptually relevant information at greatly reduced size. This means that all perceptually non-relevant information is eliminated which should facilitate information extraction. While there have been some investigations of pattern recognition in compressed domain in the past, in this paper we analyze the problem from the compression and information reduction perspective. Pattern recognition method based on optimized quantization of DCT blocks and density of blocks in regions is introduced and illustrated on the example of face detection and recognition problem.

1. INTRODUCTION

Images and video are handled nowadays to a great extent in compressed formats based on block DCT transforms. This facilitates storage and transmission. Traditional pattern recognition methods are based mostly on processing in the original picture domain, compression is not taken into account. However, lossy compression methods are highly optimized for generating description of images in which highly relevant perceptual information is preserved and all non-relevant information is eliminated. In result the number of bits for perceptual description is minimized. This is potentially very attractive from the pattern recognition point of view since it reduces redundancy from the processing. In particular, the quantized block DCT is known to minimize the number of bits while preserving perceptually important features [13]. On the other hand the block DCT operates by decomposition of local signal in frequency domain and some analogy with the operation of biological visual processing can be drawn [1].

Information extraction using DCT has been studied in the past [1]-[10] using many different techniques and combinations with other methods. While this previous work has shown that information extraction in the DCT is certainly feasible the advantages of DCT from the perceptual compression point of view were not utilized in our opinion.

In this paper we present an approach to information extraction in the DCT domain taking into account intrinsically the compression properties of the DCT. Our basic idea is that compression reduces the number of different blocks in the picture. These blocks can be grouped and classified according to their number and location. We present our ideas on the example of face detection and recognition problem. We show that quantization of DCT face images leads to specific distributions of DCT blocks. Histograms of block distribution have long tails. Certain block patterns appear often and many block patterns are rare. We show that quantization levels can be selected at which there already well-performing retrieval of face images from database is based on histograms only, without any block location information. [11]

We are also showing that quantized blocks are located in such a way that rare blocks are grouped in areas highly relevant to face information (eyes, nose, lips, etc.). Thus, to facilitate recognition one needs to identify areas with high density of blocks. In this paper we present a method for grouping of blocks and detecting areas of blocks. Measurements performed on those areas provide critical information for face pose evaluation and face recognition.

2. FEATURE DESCRIPTION USING DCT

In compression methods 8x8 DCT block transform is used. However, when higher quantization levels only the frequency components in the 4x4 areas of the DCT are nonzero. This is equivalent to 4x4 blocks DCT performed

on scaled down images. For the information extraction we thus use 4x4 DCT blocks. The 2-D DCT can be calculated directly by:

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

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

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

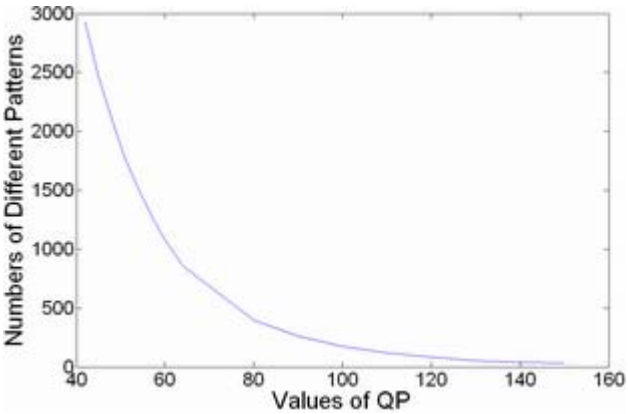

Fig.1 QP vs. Number of Patterns

In compression applications the DCT is quantized. In image compression standards quantization is preformed in very sophisticated way to optimize picture quality, in our approach high quantization levels on the equivalent 4x4 DCT block are used with scalar quantization factor QP similar to the H.264 standard [13]

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. Depending on the QP

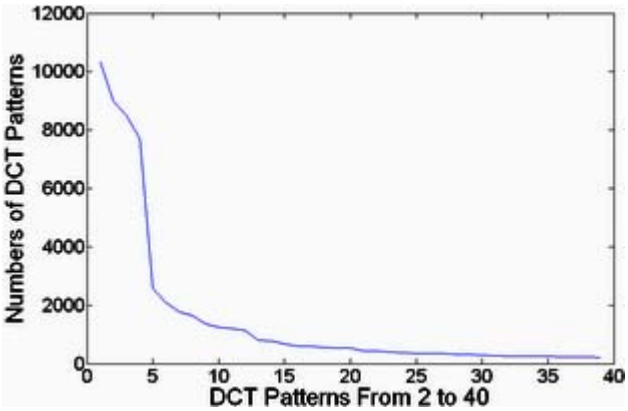

Fig.2 Probability of Patterns (QP=21)

value the number of different blocks is limited as shown in Fig. 1.

Typical block distribution for face image with specific QP factor is shown in Fig. 2. It can be seen that pattern distribution has long tail. There is limited number of patterns which appear in large quantities and significant number of patterns which appear rarely.

Basic observation used in this paper concerns location of the DCT patterns in the images. We split the DCT blocks into two sets: one for that with high quantity and one for those with low quantity. The point of splitting is not very critical. After the splitting, the face images get the following appearance shown in Fig. 3(G). It can be seen that that the set corresponding to rare patterns (black) is distributed over important face features, the set corresponding to common patterns constitute the bulk of the face.

As mentioned before, DCT block patterns can be grouped by evaluating their probabilities of occurrence. Some patterns are common in a particular set of pictures while others are not. Our research shows that, for front face only pictures which have been strongly quantized, the remaining DCT coefficients are most likely to occur at the position near to DC value. Fig. 3 (A) shows two face images. Fig. 3 (B) shows that most of the blocks are DC blocks (The blocks which do not have AC coefficients). These DC blocks are marked as white area in Fig. 3 (B).

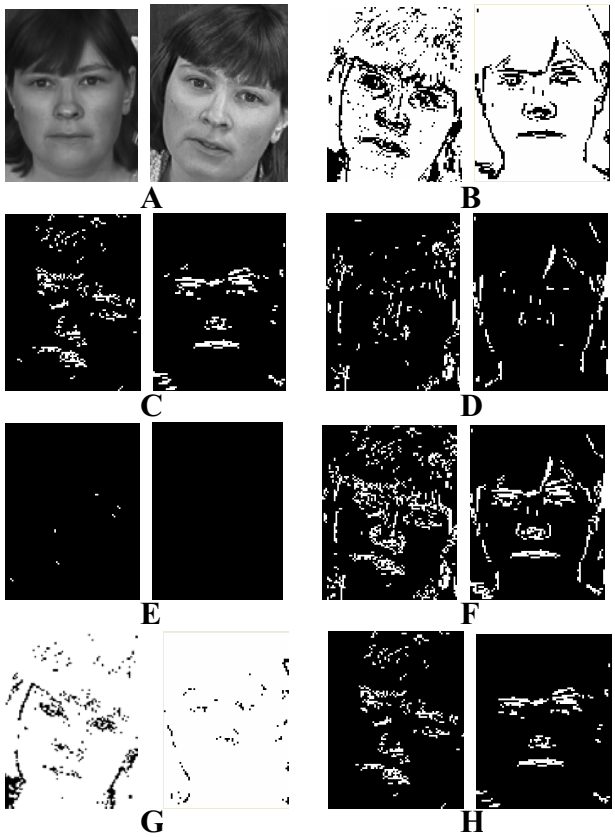

Fig.3 Rare Patterns Depict the Key Features

3. FACE REGIONS

As shown above critical face areas have very specific block distribution. We can classify the blocks further

according to the distribution of their coefficients.

Fig. 3 (C-E) show the positions of the blocks which have AC values only at the position (0, 1), (1, 0) or (1, 1) respectively. Fig. 3 (F) shows the combination of Fig. 3 (C-E). Fig. 3 (G) shows the combination of Fig. 3 (B and F). As one can see, these blocks together make a rough outline of the face. Now it is possible for us to locate the eye, nose and mouth. The areas sets are located in the following way:

1. We form 4x4 DCT transform blocks
2. Perform quantization of the DCT blocks
3. Binarize the blocks coefficients, nonzero coefficients are set to 1

*	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

Block Pattern A

*	1	0	0
0	0	0	0
0	0	0	0
0	0	0	0

Block Pattern C

*	0	0	0
1	0	0	0
0	0	0	0
0	0	0	0

Block Pattern B

*	0	0	0
0	1	0	0
0	0	0	0
0	0	0	0

Block Pattern D

Here the notation "*" stands for any DC value.

Fig.4 Most Common Blocks

4. Match each block with the specific block pattern A, B, C or D in Fig. 4, comparing only their AC coefficients. If it is matched, we set a value '1' for this block, namely, the white point; otherwise, we set a value '0' for this block, namely, the black point. Finally, we will get a 1/16 down-scaled binary image.

From these figure, we can conclude that:

1. Most area of the face and hair are DC blocks, which is corresponding to the block Pattern A.
2. Eyes, mouth and bottom line of nose are typically depicted by horizontal AC coefficients, which are corresponding to the block Pattern B. This is shown by Fig. 3(C).
3. Outline of both left and right side of face are typically depicted by vertical AC coefficients, which is corresponding to the block Pattern C. This is shown by Fig. 3(D).
4. For those faces which have been rotated, there will be more count for pattern D. This is shown by Fig. 3(E).
5. A big part of the patterns of eyes, nose and mouth do not belong to the Pattern A, B, C, and D.

4. BLOCK DENSITY MATCHING TO FIND CONNECTIVITY INFORMATION

After the main area of face has been located, locations of particular key features of the face, such as eyes, nose and mouth can be found. We call this Block Connectivity Information. By evaluating the distances and angles between eyes, nose and mouth, one can also find the pose of the face

In order to locate the position of eyes, nose and mouth, here we introduce a "Block Density Matching" method. We start from searching for eyes. As one can see in the Fig. 3(C), the areas of eye, nose and mouth are mainly consisted of white dots, and the density of white point in these areas are higher than the other parts of face. So we can locate them by evaluating the maximum density.

1. We generate a template for the eye, which is a rectangular pixel block. All the pixels are set to 1 (white point).
2. We set the pixels in the four corners to '0' (black points), to make the shape of this polygon similar to the outline of eye. To some degree, this follows the eye oval (Fig. 5).
3. Moving this template as a sliding window in a certain area, matching the area of simple AND operation, we can roughly locate the location of the eyes.
4. The template can be adapted to the size of the face by changing its size. This can slightly improve the accuracy of searching result, while introduce more computation complexity.

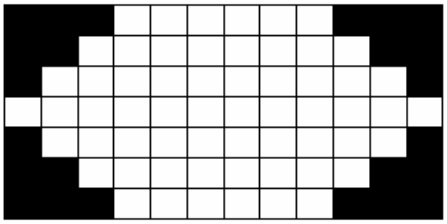

Fig.5 Templates Used for Matching

After we have located the position of eyes, we would next try to locate the position of nose. Here we could use some a priori knowledge to deduce the searching area. For example, for the non-rotated front face, if the center position of left eye, right eye, nose and mouth are respectively at (X_1, Y_1), (X_2, Y_2), (X_3, Y_3) and (X_4, Y_4). Then one can easily draw the conclusions below:

1. $X_2 > X_3$, $X_4 > X_1$; $X_3 \approx X_4$
2. $Y_1 \approx Y_2$; $Y_4 < Y_3 < Y_1, Y_2$
3. X_3 is near the value of $(X_1+X_2)/2$
4. Y_3 is near the value of $(Y_1+Y_2)/2 + (X_2-X_1)x2/3$
5. X_4 is near the value of X_3
6. Y_4 is near the value of $(Y_1+Y_2)/2 + (X_2-X_1)$

After these rough searching areas have been determined, we can perform full-pixel matching, which is similar to previous one. One can finally find out the value of (X_3,

Y_3) and (X_4, Y_4). The template used for searching nose is the same or a little larger width of the one used for eyes; while the template for mouth has 1.5 to 2 times width of the one used for eyes.

Furthermore, if the $X_3 \approx (X_1+X_2)/2$, then we can deduce that the face is at the front position; otherwise, the face maybe horizontally turns to one side. If X_3 is more closer to X_2, then the face turns to left side; if not, the face turns to right side.

This relationship between these key face features, i.e. Block Connectivity Information, is actually what we are looking for. Fig.6 shows an example of the Block Density Matching. The searched positions of eyes and nose notated by two gray lines, which are the top and bottom lines of matching template.

However, it would be more difficult for the case of rotated face. For the rotated face, $Y_1 \neq Y_2$, the two eyes are not in a same horizontal line, but in a diagonal line. The position of nose and mouth need more calculation, but the basic relation remains the same. This is also our future research

From our experiment result we found that, the hair is a major distraction factor. To achieve more accurate result, we should also remove the hair. As we mentioned before, the area of hair is mainly composed of DC block patterns. And the DC value of hair is clearly different from the DC value of face. Therefore, we set a threshold value to filter the compressed image; each DC value which is below the threshold would be set "1". Now the eyes can be easily discriminated. Fig.3 (H) is an example based on Fig.3(C).

5. EXPERIMENTAL SYSTEM AND RESULTS

For experiments we used the Georgia Tech Face (GTF) Database [12]. The experiment result shows that: for most of the front positioned face, we can locate their key features. Figure 6 shows the example of person s26 in the database. The position of eyes, nose and the both side of face are noted by a pair of white lines.

6. CONCLUSION

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 of "Block Density Matching" is illustrated based on quantized 4x4 DCT blocks of face database images. By matching the template through pattern indexed images, good results of key features recognition and locating are obtained. The method is computationally efficient and can be used directly for information extraction from compressed video. Further research is needed for dealing with the rotated and turned face images.

Fig.6 Block Density Matching Result

REFERENCES

[1] Yu Zhong and Anil K. Jain, "Object localization using color, texture and shape", Pattern Recognition, Vol.33, No.4, pp.671–684, Apr. 2000

[2] Richard E. Frye, Robert S. Ledley, "Texture discrimination using discrete cosine transformation shift-insensitive (DCTSIS) descriptors", Pattern Recognition, Vol.33, No.10, pp.1585-1598, October 2000

[3] S. Eickeler, S. Muller and G. Rigoll, "Recognition of JPEG Compressed Face Images Based on Statistical Methods", Image and Vision Computing, Vol. 18, No. 4, pp. 279-287, 2000

[4] Ramasubramanian, D. and Y.V. Venkatesh, "Encoding and recognition of faces based on the human visual model and DCT", Pattern Recognition, Vol.34, No.12, pp. 2447-2458, December 2001

[5] Jie Wei, "Image segmentation based on situational DCT descriptors", Pattern Recognition Letters,Vol 23, No.1-3, pp. 295 - 302, January 2002

[6] Jiang J., Armstrong A. and Feng GC "Direct content access and extraction from JPEG compressed images", Pattern Recognition, Vol.35, No.11, pp.2511-2519, November 2002

[7] C. Sanderson and K. K. Paliwal, "Fast Features for Face Authentication Under Illumination Direction Changes", Pattern Recognition Letters, Vol. 24, No.14, pp.2409 - 2419, 2003.

[8] Min-Sub Kim, Daijin Kim and Sang-Youn Lee, "Face recognition using the embedded HMM with second-order block-specific observations, Pattern Recognition, Vol. 36, No. 11, pp.2723-2735, November 2003

[9] GuocanFeng,Jianmin Jiang, "JPEG compressed image retrieval via statistical features", Pattern Recognition, Vol.36, No. 4, pp. 977-985, April 2003

[10] C.Sanderson and K.K.Paliwal, "Features for Robust Face Based Identity Verification". Signal Processing Vol.83, No.5, pp.931-940, May 2003

[11] Koji Kotani, Chen Qiu, and Tadahiro Ohmi, "Face Recognition Using Vector Quantization Histogram Method," in *International Conference on Image Processing*, II-105, Sep. 2002.

[12] Georgia Tech Face Database, Available: ftp://ftp.ee.gatech.edu/pub/users/hayes/facedb/.

[13] Joint Video Team(JVT) of ISO/IEC MPEG&ITU-T VCEG, "Draft ITU-T Recommendation and Final Draft International Standardd Joint Video Specification", Document JVT-G050,March 2003