A STUDY ON THE EFFECT OF ROI MASKS ON FACE RECOGNITION SYSTEM USING DIGITAL RECORDER

Mitsuhiro Fujita, Takahiro Yoshida, and Seiichiro Hangai

Department of Electrical Engineering, Tokyo University of Science
1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan

ABSTRACT

In recent years, biometrics authentication has been widely used to realize high security in airports, office buildings and so on. In public area, surveillance cameras, which are combined with a high efficient coding recorder such as MPEG2 or MPEG4 recorder, are set to take face images of pedestrians and car license plates. However, high compression for efficient recording leads to degradation of details in face and characters on car license plates. Consequently, it is important that region of interest (ROI) is detected correctly and recorded with high quality image.

Although there are a number of proposals [1,2] concerning the detection of face areas and license plates, the relationship between ROI masks and face recognition rate or character recognition rate is not cleared.

In this work, we investigate the recognition rates for face authentication with compressed face images under limited storage size. Several kinds of ROI mask are compared, and face recognition rates with respect to ROIs are shown.

1. INTRODUCTION

Surveillance systems have been used in various areas, such as airports, office buildings and toll gates. In the old system, moving pictures were recorded in analog on magnetic tapes, and important scenes were manually searched. Although the SNR was not so good, there was a possibility of extracting features of face and shape of characters by posteriori processing. In recent years, digital video cameras are widely used and moving pictures are recorded on HDD with high efficient coding scheme in surveillance systems. Though the density of HDD is becoming high without increasing cost, customer's requirements make it difficult to record moving pictures on HDD with high quality. This is because the surveillance camera works 24 hours a day, 7days a week, and 365days a year. This means that the maintenance period of one week requires 168GB storage, one month requires 5TB storage, and one year requires 60TB storage, if the picture quality of MPEG2 is similar to that of analog video recorder. Therefore, in the worst case after the picture quality is lowered by the reason of storage size, it is impossible to identify objects in the movie afterward. Though the size of the movie can be decreased by the moving detection [3], it is unsuitable for the use of public facilities. Actually, moving pictures are temporarily thinned out and recorded as still pictures.

In order to decrease the size of the movie without removing a possibility of extracting features, essential areas in the movie should be encoded with high quality, and the other areas with low quality. Thereby, for example, if a face area is encoded with high quality, it helps computers or us to recognize the face in the image. The ROI encoding will become one of solution for the above problem. As for JPEG, R.L.V. Hue et al. [4] developed JPEG ROI encoding and showed normalized match rates for the sizes of token image files. In JPEG 2000, the ROI encoding is adopted by default [5], and then any areas in an image are specified if encoders support the ROI encoding. A. Leontaris et al. [6] also proposed a movie encoding with ROI for the purpose of reducing the size of the encoded movie.

In this work, we investigate the effect of interested areas for face recognition with still images by using the ROI encoding in JPEG 2000. After face image is encoded with several ROI masks, SVM (Support Vector Machine) [7] classifier trained with non-compressed face images identifies face in the decoded image. And PCA (Principal Component Analysis) is applied for creating feature vectors used to train the SVM. As a result, we show the effectiveness of ROI mask for face recognition. Eventually, this work will be combined with the face detection [8], and we will build the system of face authentication for pedestrians.

2. AUTHENTICATION SYSTEM WITH THE ROI ENCODING

Figure 1 shows an overview of the proposed authentication system. In this system, the processes are as follows:

(1) By referring row image data from surveillance cameras, the ROI Detector defines ROI.

(2) Based on the ROI information, the images are encoded and saved as a JPEG2000 still picture.

(3) In face identification, compressed movies are decoded. Then, the identifier can selectively search areas where more bits should be allocated.

In process (3), because the positions of interesting areas such as faces or car license plates are easily estimated by the use of ROI encoding, this method has the advantage in cost of calculation.

3. METHOD OF FACE RECOGNITION

3.1. Experimental flow

The experimental procedure is shown in Figure 2. This consists of two procedures, i.e., the training procedure and the identifying procedure.

For the training procedure, from scanned face images, face areas are cropped. Then, these areas are processed in order to adjust illumination conditions. After that, in training a model, by adapting PCA to processed areas, eigenvectors and eigenvalues of these areas are provided. Finally, the learning model is trained with feature vectors linear-transformed from the processed areas.

In the identifying procedure, ROI masks are created from the scanned images. Next, by using JPEG 2000 encoder, the images are encoded with ROI masks and the encoded images are decoded. Then, just like the training procedure, face areas in decoded images are cropped and processed. Conclusively, the model identifies the face areas by determining the feature vectors linear-transformed from the areas with the eigenvectors.

3.2. Test images

We use a studio portrait taken in photo shops. The color photograph is for a passport, and the size is 3.5x4.5 cm. It is scanned by the scanner, EPSON GT-8700, with 400 dpi. The size of the scanned image is approximately 551x709 pixels. As shown in Figure 3, the image is snipped with a ratio of 3:4 as large as possible. Then, it is reduced to 480x640 by averaging pixels. In determining the scan resolution and image size, we refer the token face image type in [9]. After that, the face area in the image is determined manually with a form of square. In Figure 3, its area is clipped with the rectangle. In this work, we prepare 29 images obtained from 29 persons.

3.3. ROI masks

By using the scanned image and the information of the face area, ROI masks are created as shown in Figure 4 (1)-(6). Previously, the image is transferred to grayscale image. The mask (1) contains all parts of the face area. The mask (2) and (3) contain the upper and lower of the face area respectively. The mask (4) contains inscribed circle of the face area. The mask (5) contains the area that the horizontal radius of the circle in the mask (4) is reduced with 50% area of the face. The mask (6) contains the area that the edge area of the face area is binarized with 50% area of the face.

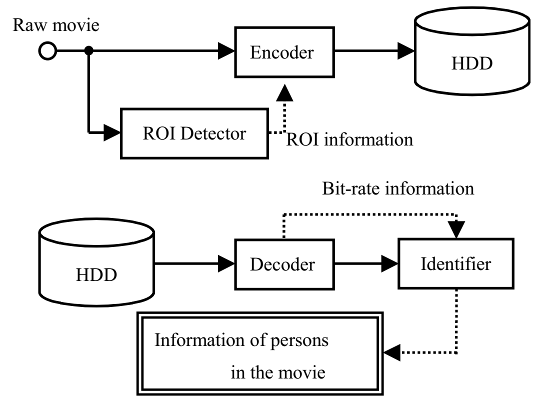


Figure 1: The overview of proposed authentication system

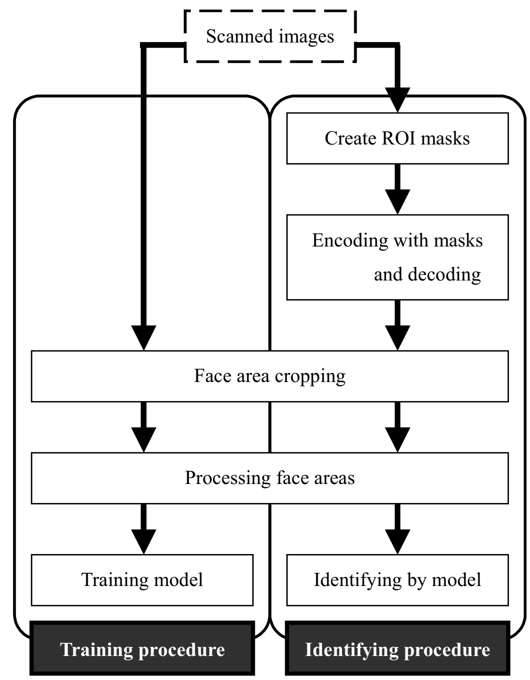


Figure 2: The procedure of face recognition

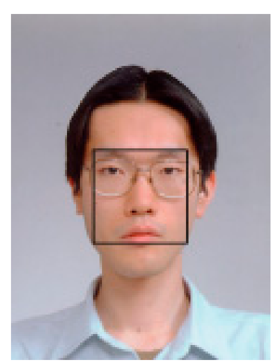


Figure 3: Scanned image with face area

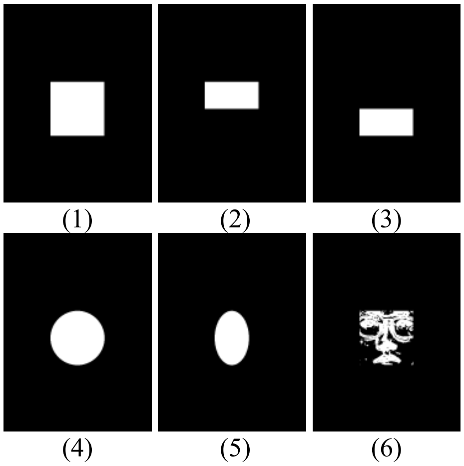


Figure 4: ROI masks

3.4. Encoding with JPEG 2000

The scanned image is encoded with JPEG 2000. We use JJ2000 [10] as an encoder. The image is encoded without masks or with six masks shown in Figure 4. Simultaneously, bit rates are given to the encoder. As a result, 7 encoded images are created with respect to each bit rate.

3.5. Cropping and processing face areas

Face area is cropped from non-compressed image for training the model as shown in Figure 2. In the face identification, face area cropped from decoded image is used. The areas are transferred to grayscale image and reduced to 32x32 by averaging pixels. And the Gaussian smoothing filter and the histogram equalization are applied to the areas.

3.6. Training identification model

From processed face areas, feature vectors are created and fed through the identification model. In order to reduce the dimensions of feature vectors, the PCA is used. Then, the vectors are used to train the SVM.

3.6.1. Principal Component Analysis

PCA is equal to solving the eigenvalue problem. In particular, principal components from face images are called as eigenface, and the method is often used for face recognition and face detection [11-13]. The first component is obtained from the eigenvector corresponding to the maximum eigenvalue, and this component has the maximum variance. Therefore, for identifying face images, each face is easily classified with low dimensions of feature vectors.

In this study, the data from the face area has 1024 dimensions. One dimension corresponds to one pixel, and one pixel ranges from 0 to 255. In the first step, each value in the image is normalized from 0 to 1. In the second step, each image is assumed as one-dimensional vector, and the eigenvalue problem with the covariance matrix from these vectors is solved. In the final step, the new vector is obtained by using linear transformation from the original vectors to the eigenvectors. The dimensions of new vectors are 10, 20, 100, and 1024 (equal to original vectors).

3.6.2. Support Vector Machines

SVMs are binary classifiers, and these have been used among face detection and recognition [7].

We have implemented SVMs by using LIBSVM [14]. Since a face image is identified from registered face images, the implemented SVM is the *n*-class classifier (let *n* be the number of registered persons). Each person corresponds to one face image, and the SVM is trained with *n* vectors obtained above section. For the parameters of SVM, the installed kernel is RBF (Radial Basis Function), and all of same parameters are given to the SVM in every case.

3.7. Identifying faces by the model

For identifying face images, the feature vector is created by transforming each processed image with eigenvectors. The vector is determined by the SVM, and if the obtained number is equal to true number of the person, we regard that the identification of the person is successful.

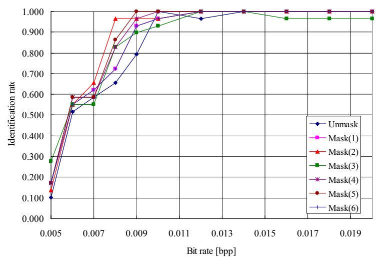
4. EXPERIMENTAL RESULTS

Figure 5 indicates line graphs that represent the relationship of bit rates and successful identification rates with respect to vectors of 10, 20, and 1024 dimensions respectively. The range of bit rates is from 0.005 bpp (bits per pixel) to 0.020 bpp. We have confirmed that the result of 0.020 bpp or more values is similar to one of 0.020 bpp in JJ2000.

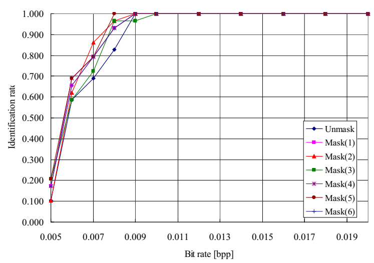
5. DISCUSSION

For the result of 10 dimensions, the unmasked identification rate is the lowest. In Figure 5 (a), between 0.007 bpp and 0.008 bpp, the result of mask (2) shows particularly high identification rate. This indicates that the area around eyes is important in low bit rates. However, in this dimension, each identification rate has variations compared with the other dimensions. Additionally, although the identification rate of the mask (3) goes up to 1.000 at 0.01bpp, it falls down at 0.016 bpp. It is found that the mask does not contribute to the improvement of the identification even if the incapable areas such as the lower area of face are compressed with high bit rate.

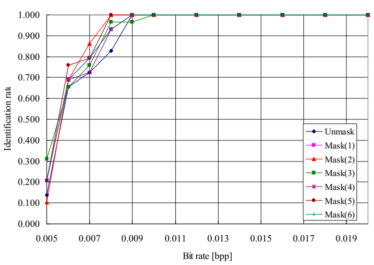
For the results of 20 dimensions in Figure 5 (b), each identification rate achieves 1.000 in lower bit rates than that of the lower dimension. In particular, the rates of the mask (2) and the mask (5) increase rapidly. In contract, the mask



(a) 10 dimensions



(b) 20 dimensions



(c) 1024 dimensions

Figure 5: The relationship of bit rates and identification rates

(3) increases slowly. Additionally, the results are almost similar to these of 100 dimensions.

For the result of 1024 dimensions, the amount of information for the identification can be fully used. The result denotes the same tendency of 20 or 100 dimensions. Moreover, the bit rates where achieve 1.000 are least among 4 kinds of dimensions.

From above experimental results, the shapes of effective ROI mask for face identification are the mask (2) and the mask (5), which cover eyes and important parts of a face. For the mask (6), though the mask covers the edge that represents face features and has the same area of the mask (2) or (5), the result of the mask (6) becomes worse than that of the mask (2) or (5). The reason is that the form of the ROI mask is too complex to compress with high quality. For the mask (1) or (4), the results are not so bad. However, enough bit rates are not allocated to masked areas because of small bit rates given.

6. CONCLUSION

In this paper, we have compared the performance of face identification rate with various kinds of ROI masks based on JPEG2000. From the result, the mask that covers eyes and important parts of face contributes both conservation of the identification rate and the low storage size.

In future work, we will investigate the optimum mask derived from the image for face identification.

7. REFERENCES

[1] M.-H.Yang,D.Kriegman, and N.Ahuja, “Detecting faces in images: a survey,” *Pattern Analysis and Machine Intelligence*, Vol.24, No.1, pp.34-58, 2002.

[2] S.Yohimori,Y.Mitsukura,M.Fukumi,N.Akamatsu, and N.Pedrycz, “License Plate Detection System by Using Threshold Function and Improved Template Matching Method,” *NAFIPS '04*, Vol.1, pp.357-362, 2004.

[3] T.Suzuki, H.Sekino, N.Saeki, and K.Iwata, “Sensor Detection Method in AI Camera,” *FUJITSU GENERAL TECHNICAL JOURNAL*, No.2, 1999.

[4] R.L.V.Hsu and P.Griffin, “JPEG Region of Interest Compression for Face Recognition,” IDENTIX Document # RDNJ-04-0102, http://www.identix.com/research/JROI.pdf, 2004.

[5] T.Acharya and P.S.Tsai, *JPEG2000 Standard for Image Compression*, John Wiley & Sons, Inc.

[6] A.Leontaris and P.C.Cosman, “Region-of-Interest Video Compression with a Composite and a Long-Term Frame,” *IASTED International Conference on Computers, Graphics and Imaging*, pp.435-440, 2004.

[7] N.Cristianini and J.Shawe-Taylor, *An Introduction to Support Vector Machines and other kernel-based learning methods*, Cambridge University Press, 2000.

[8] M.Fujita, T.Yoshida, and S.Hangai, “Face Area Localization using Support Vector Machines based on Color Spaces,” *World Multi-Conference on Systemics, Cybernetics and Informatics*, Vol.5, pp.326-329, 2005.

[9] ISO/IEC CD 19794-5, “Biometric Data Interchange Formats --- Part 5: Face Image Data,” http://www.jtc1.org/FTP/Public/SC 37/DOCREG/37N0342.pdf.

[10] “JJ2000 Public Homepage,” http://jj2000.epfl.ch/index.html.

[11] M.Turk and A.Pentland, “Eigenfaces for Recognition,” *Journal of Cognitive Neuroscience*, Vol.3, No.1, pp.71-86, 1994.

[12] Y.Li, S.Gong, J.Sherrah, and H.Liddell, “Multi-view Face Detection Using Support Vector Machines and Eigenspace Modelling,” *Knowledge-Based Intelligent Engineering System & Allied Technologies*, pp.241-244, 2000.

[13] Vlad Popovici, and Jean-Philippe Thiran, “Face Detection Using an SVM Trained in Eigenfaces Space,” *Audio- and Video-based Biometric Person Authentication*, pp.190-198, 2003.

[14] C.C.Chang and C.J.Lin, “LIBSVM: A Library for Support Vector Machines,” http://www.csie.ntu.edu.tw/~cjlin/libsvm.