MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan

8-10

A New Scheme for Image Recognition Using Higher-Order Local Autocorrelation and Factor Analysis

Naoyuki Nomoto†, Yusuke Shinohara†, Takayoshi Shiraki†, Takumi Kobayashi†, Nobuyuki Otsu† †††

†The University of Tokyo †††AIST
Tokyo, Japan Tukuba, Japan

shiraki, takumi, otsu @isi.imi.i.u-tokyo.ac.jp

Abstract

This paper proposes a new scheme for multipurpose image recognition based on Higher-order Local Auto-Correlation (HLAC) features and factor analysis. First, HLAC features, which are inherently invariant under translation, computationally inexpensive, and additive, are extracted from the input images. Second, factor analysis is applied to the feature vectors so as to decompose the feature vectors as combinations of factors leant through supervised training examples. After the factorization, the input image is recognized by using the factor scores obtained through the least squares method. Experimental results show that the proposed method effectively enables the system to recognize images by acquiring effective factors that represent each object in the images without any need for segmentation or locating objects.

1. Introduction

Image recognition has become of great interest over the past decades because of its potential applications in many fields, such as optical character recognition (OCR), identity authentication, human-computer interfacing, and surveillance. Numerous algorithms have been proposed for image recognition (for instance, a detailed survey on facial expression recognition is given in [3]).

In this paper, we present a new scheme for image recognition for various purposes. Our method consists of two stages. In the first stage, Higher-order Local Auto-Correlation (HLAC) features are extracted. The HLAC features proposed by Otsu have been successfully applied to face recognition and many other pattern recognition problems [9, 10, 6, 5, 4], and are inherently shift-invariant, additive, and computationally inexpensive. In the second stage, effective features or factors to represent each object in the images are learnt by using factor analysis. The HLAC feature vector extracted in the first stage is then decomposed by using these factor vectors to count the numbers of each object. To obtain from detailed to rough information of the provided images, HLAC features are extracted from each pyramidal image. The importance of multi-resolution (multi-channel) processing has often been pointed out, and this type of processing has often been used to increase recognition rates [2, 5, 6, 8]. Edge images are also used to obtain robustness regarding variations in the lighting conditions. The solutions from each resolution of each type of image (gray scale and edge) are integrated in the final stage to obtain the final solution of the system.

To verify the effectiveness of this scheme, we applied the proposed scheme to two types of recognition task. One was a task requiring the system to count objects of various shapes on a newspaper; the other required face recognition.

2. Feature Extraction

2.1. HLAC features

Based on [6, 10], we use HLAC features. Let an image plane be denoted as P. Higher-order autocorrelation functions in general are then defined by

$$^{N}(\boldsymbol{a}_1 \quad \boldsymbol{a}_N) = \int_P \ (\boldsymbol{r}) \ (\boldsymbol{r}+\boldsymbol{a}_1) \quad (\boldsymbol{r}+\boldsymbol{a}_N) \ \boldsymbol{r}$$

where N denotes the order of the autocorrelation functions, $\boldsymbol{r}$ is the image coordinate vector, and $\boldsymbol{a}_i$ are the displacement vectors. $(\boldsymbol{r})$ denotes the gray-level at position $\boldsymbol{r}$.

The number of autocorrelation functions obtained by combining the displacements over the region P is enormous. Here, we reduce this number to enable practical application. First, we restrict the order N to the second order ($N = 0 \ 1 \ 2$). Then, we also restrict the range of displacements to within a local 3×3 region, whose center is the reference point (Figure 1), because the correlation within a local region is much higher than the correlation between distant points. By eliminating displacements that are equivalent because of an even shift, we reduce the number of the displacement patterns to 25. Figure 2 shows the 25 types of local displacement patterns, where the symbol * represents "don't care". Thus the HLAC features are obtained by scanning the image over P with the 25 local 3×3 masks and by computing the sums of the products of the gray values of the pixels corresponding to "1". The HLAC features are obviously additive for isolated objects on P, which is an important and desirable characteristic for applying factor analysis. Also, HLAC features are shift-invariant, which makes the system robust to changes in the position of objects within an image.

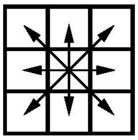

Figure 1: Local 3 × 3 region and the reference point

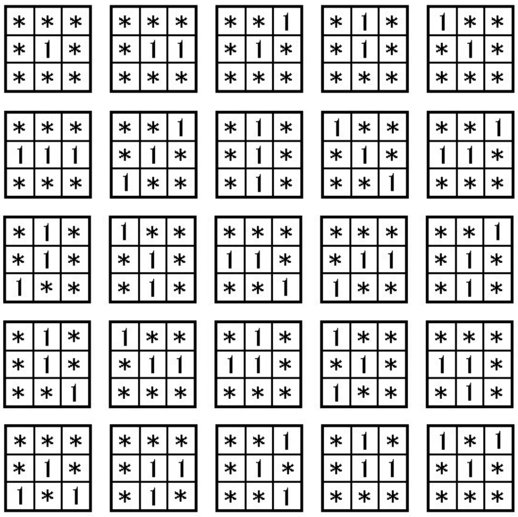

Figure 2: 25 types of local displacement patterns used for computing HLAC features, where the symbol * represents "don't care".

2.2. Feature integration

The primitive HLAC features computed from the image of the highest resolution include only local and too detailed information. Often images can be more easily recognized with rough resolution. Thus, we construct a pyramidal image data structure which gives a set of images with different resolutions ranging from high to low [1]. We also use edge images and the pyramid structure of edge image to gain robustness regarding illumination variation. Figure 3 shows the feature extraction scheme. A set of the HLAC features extracted from each image in the pyramidal structure includes information ranging from detailed to rough.

The solutions (recognitions) from each pyramidal layer are integrated to obtain the final solution. This integration is done through majority decision with proportional weight assigned to the learning rate.

3. Factor Analysis of HLAC Features

A set of 25-dimensional HLAC feature vectors $\boldsymbol{z}_i\ {}_{i=1}^{N}$ is extracted from an image or edge image G_i at each resolution layer. Suppose that each image consists of some of common isolated image regions (objects), whose HLAC feature vectors are $\boldsymbol{f}_j$ ($= 1$) and we call $\boldsymbol{f}_j$ the factor vectors. Since HLAC features are additive, $\boldsymbol{z}_i$ can be approximated as a linear combination of $\boldsymbol{f}_j$ as follows.

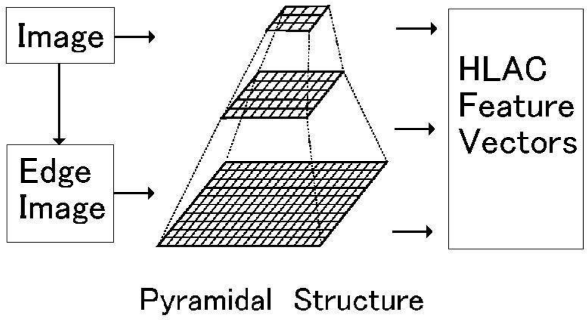

Figure 3: HLAC feature extraction

$$
\begin{aligned}
\boldsymbol{z}_i &= a_{i1}\boldsymbol{f}_1 + \quad + a_{im}\boldsymbol{f}_m + \boldsymbol{e}_i \\
&= [\boldsymbol{f}_1 \quad \boldsymbol{f}_m][a_{i1} \quad a_{im}]^T + \boldsymbol{e}_i \\
&\triangleq F^T\boldsymbol{a}_i + \boldsymbol{e}_i
\end{aligned} \tag{1}
$$

where $F = [\boldsymbol{f}_1 \quad \boldsymbol{f}_m]^T$ is the unknown factor matrix ($\times 25$), $\boldsymbol{a}_i$ is the coefficient (score) vector ($\times 1$) whose components denote the number of each factor corresponding to each object, and $\boldsymbol{e}_i$ is the residual error.

We assume that the HLAC features $\boldsymbol{z}_i$ of image G_i and are given, and $\boldsymbol{a}_i$ are given as answers in the supervised learning process. The mean square error can then be written as

$$
\begin{aligned}
{}^2[F] &= \mathop{\mathrm{E}}_i \|F^T\boldsymbol{a}_i \quad \boldsymbol{z}_i\|^2 \\
&= \ r(F^T R_{aa} F) \quad 2\ r(F^T R_{az}) + \mathop{\mathrm{E}}_i [\boldsymbol{z}_i{}^T\boldsymbol{z}_i]
\end{aligned} \tag{2}
$$

where $\mathop{\mathrm{E}}_i$ denotes $\sum_{i=1}^{N} /N$, $R_{aa} \triangleq \mathop{\mathrm{E}}_i [\boldsymbol{a}_i\boldsymbol{a}_i{}^T]$ denotes the autocorrelation matrix of $\boldsymbol{a}_i$, and $R_{az} \triangleq \mathop{\mathrm{E}}_i [\boldsymbol{a}_i\boldsymbol{z}_i{}^T]$ denotes the cross-correlation matrix between $\boldsymbol{a}_i$ and $\boldsymbol{z}_i$.

The optimal factor matrix F that minimizes Eq. (2) is obtained by solving the equation

$$
\frac{\partial\ {}^2[F]}{\partial F} = 2R_{aa}F \quad 2R_{az} = O
$$

Therefore, we obtain

$$
F = R_{aa}^{\ 1} R_{az} \tag{3}
$$

The factor analysis is conducted on each resolution layer.

4. Recognition

Taking advantage of the factor analysis, we can use the least squares method (LSM) to recognize an image.

Let $\boldsymbol{z}$ be the HLAC feature vector of the image to be recognized. Factor matrix F is given by the solution of Eq. (3) in the learning process. The optimum coefficient vector $\boldsymbol{a}$ that indicates the numbers of corresponding factors (objects) in the image is determined so as to minimize the square error between both sides of the following equation in an ideal case.

$$\boldsymbol{z} = F^T \boldsymbol{a} \tag{4}$$

The square error is

$$\begin{aligned} {}^2[\boldsymbol{a}] &= \|F^T \boldsymbol{a} \quad \boldsymbol{z}\|^2 \\ &= \boldsymbol{a}^T F F^T \boldsymbol{a} \quad 2\boldsymbol{a}^T F \boldsymbol{z} + \boldsymbol{z}^T \boldsymbol{z} \end{aligned} \tag{5}$$

By solving the equation

$$\frac{\partial \; {}^2[\boldsymbol{a}]}{\partial \boldsymbol{a}} = 2FF^T\boldsymbol{a} \quad 2F\boldsymbol{z} = 0$$

we obtain the solution

$$\boldsymbol{a} = (FF^T) \;\; {}^1 F\boldsymbol{z} \tag{6}$$

provided that FF^T is non-singular.

5. Experiments

To evaluate the performance of the proposed method, we conducted two types of experiment. Recognition targets were 2D shapes and human faces.

5.1. Counting objects on a newspaper

In this experiment, the system was required to count the number of various shapes (circles, squares, or triangles) on a newspaper. Figure 4 shows three samples randomly selected from the dataset.

We used 270 images in the experiment and the size of each original 8-bit gray scale image was 400 $\times$ 400 pixels. HLAC features were extracted from each image in a pyramidal structure with 37 layers (3 $\times$ 3, 4 $\times$ 4, , 9 $\times$ 9, 10 $\times$ 10, 20 $\times$ 20, , 300 $\times$ 300). We acquired a series of images from different objects and used the hold-out method (*training set* (190 images) *test set* (80 images) =).

From each layer, we obtain coefficients which correspond to the number of each object. Since the number of each object is an integer, coefficients are rounded before majority decision.

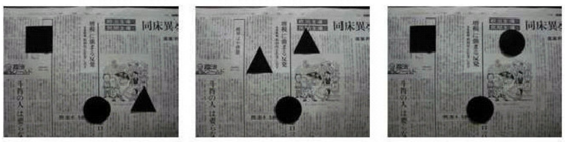

Figure 4: Examples of images. Simple objects (circles, squares, or triangles) laid on a newspaper.

Table 1 shows the recognition rates for the final solution of the system. The final solution was obtained through majority decision. The training set was obtained randomly for each trial and we performed this experiment 10 times; the results shown in the table are the average of our results for each trial. The results show that the proposed method was able to consistently recognize simple shapes on a complicated background without any *segmentation* of objects.

Table 1: Recognition rates of experiment 1

shape	circle	square	triangle
Rate	100%	97.5%	100%

5.2. Face recognition

To further test the performance of our system to a more realistic and difficult recognition task, we tested it using the JAFFE facial expression database [7] which contains posed emotional facial expression images from 10 Japanese female subjects (seven different emotion displays) as shown in Figure 5. We used 193 images in this experiment (nine people with seven basic facial expressions: a neutral face (left-most), followed by happiness, sadness, surprise, anger, disgust, and fear). The size of each original 8-bit gray scale image was 256 $\times$ 256 pixels. As in experiment 1, HLAC features were extracted from each image in a pyramidal structure with 37 layers (3 $\times$ 3, 4 $\times$ 4, , 9 $\times$ 9, 10 $\times$ 10, 20 $\times$ 20, , 300 $\times$ 300).

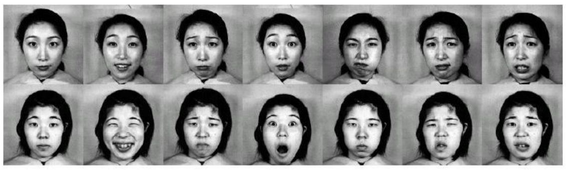

Figure 5: Sample images from the JAFFE facial expression database [7]. Note the slight variations in the head position, scale, and rotation.

In this experiment, we assumed that each image consisted of 64 factor vectors (9 for each person's neutral face factor (identity factor), 54 for each person's expression factors other than neutral face, and one for the common background).

First, factor anaysis is employed to obtain coefficients from each layer. Second, the person is identified and expression factor vector is obtained by subtracting her identity factor vector from HLAC feature vector. Third, we obtain coefficients of expression factors by the least square solution (6).

We used the same testing paradigm as Zhang et al. [11]. The entire set of images was randomly divided into 10 segments. Then, the system was trained using nine of these segments. After training, the system was tested using the remaining segment. The procedure was repeated 10 times while changing the segment left for the test phase. The recognition rate was obtained by averaging the recognition rates for these 10 experiments. Table 2 shows the recognition rates for the final solution of the system. The result indicates the effectiveness of the proposed scheme even for such a difficult recognition task.

Table 2: Recognition rates of experiment 2 : Facial expression (Exp.) and identity recognition (Id.)

Id.	Exp.	Exp.+Id.
98.9%	80.8%	79.7%

6. Conclusion

We have proposed a new scheme for image recognition which is based on HLAC features and factor analysis. Experimental results show that this method can be effectively applied to several kinds of tasks, such as counting objects on images or face recognition. It is noticed that our new scheme based on HLAC features and factor analysis eliminates any need for segmenting or locating objects in images. Other recognition tasks can be achieved through similar approaches.

Our future research will focus on analyzing ways to combine these feature extraction methods with other classifiers and determining the applicability and limitations of the proposed image-recognition method. We also plan to extend this method to other types of task relating to motion recognition, such as gesture recognition.

Acknowledgements

The authors thank Dr. Michael Lyons for providing us with the JAFFE database.

References

[1] D. H. Ballard and C. M. Brown. *Computer Vision*. Prentice-Hall, 1982.

[2] A. C. Bovik, M. Clark, and W. S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(1), 1990.

[3] B. Fasel and J. Luettin. Automatic facial expression analysis: a survey. *Pattern Recognition*, pages 259–275, 2003.

[4] F. Goudail, E. Lange, T. Iwamoto, K. Kyuma, and N. Otsu. Face recognition system using local autocorrelations and multiscale integration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(10):1024–1028, 1996.

[5] T. Kurita and N. Otsu. Texture classification by higher order local autocorrelation features. *Proc. of Asian Conf. on Computer Vision*, pages 175–178, 1993.

[6] T. Kurita, N. Otsu, and T. Sato. A face recognition method using higher order local autocorrelation and multivariate analysis. *Proc. of The 11th IAPR Inter. Conf. on Pattern Recognition*, 2:213–216, 1992.

[7] M. Lyons and S. Akamatsu. Coding Facial Expressions with Gabor Wavelets. *Proc. of 3rd IEEE Inter. Conf. on Automatic Face and Gesture Recognition*, pages 200–205, 1998.

[8] B. S. Manjunath and R. Chellappa. A unified approach to boundary perception: edges, textures, and illusory contours. *IEEE Trans. on Neural Networks*, 4(1), 1993.

[9] N. Otsu. Mathematical studies on feature extraction in pattern recognition. *Researches of the ETL*, 818, 1981. (In Japanese).

[10] N. Otsu and T. Kurita. A new scheme for practical flexible and intelligent vision systems. *Proc. of IAPR Workshop on Computer Vision*, pages 431–435, 1988.

[11] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu. Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron. *Proc. of Third IEEE Conf. on Face and Gesture Recognition*, pages 454–459, 1998.