Facial Expression Recognition Based on Local Binary Patterns[1]

X. Feng[a], M. Pietikäinen[b], and A. Hadid[b]

[a] *Northwestern Polytechnic University, Xi'an, 710072 China*

[b] *Machine Vision Group, Infotech Oulu and Department of Electrical and Information Engineering, University of Oulu, P.O. Box 4500, FIN-90014 Finland*

e-mail: fengxiao@nwpu.edu.cn; {mkp,hadid}@ee.oulu.fi

Abstract—In this paper, a novel approach to automatic facial expression recognition from static images is proposed. The face area is first divided automatically into small regions, from which the local binary pattern (LBP) histograms are extracted and concatenated into a single feature histogram, efficiently representing facial expressions—anger, disgust, fear, happiness, sadness, surprise, and neutral. Then, a linear programming (LP) technique is used to classify the seven facial expressions. Experimental results demonstrate an average expression recognition accuracy of 93.8% on the JAFFE database, which outperforms the rate of all other reported methods on the same database.

DOI: 10.1134/S1054661807040190

1. INTRODUCTION

The wide range of applications in human-computer interaction, telecommunication, and psychological research make facial expression analysis and recognition an active research topic. Approaches to facial expression analysis from both static images and image sequences have been proposed in the literature [1, 2].

Facial expression recognition from static images is a more challenging problem than from image sequences owing to the fact that less information for expression actions is available. There is a relatively small amount of work on it. However, information in a single image is sometimes enough for expression recognition, and in many applications, it is also useful to recognize a single image's facial expression.

In general, automatic facial expression recognition involves three main steps [1, 2]: face detection, feature extraction, and facial expression classification. In this paper, we focus only on feature extraction and expression classification.

There are mainly two kinds of face representations for facial expression analysis [1–3]: holistic template-based methods, for example, [4], and geometric feature-based methods, for example, [5]. In a holistic system, the template can be a pixel image or a feature vector obtained after processing the face images as a whole. In a geometric feature-based system, shape and location of facial components (including mouth, eyes, brows, nose, etc.) are extracted to form a feature vector. Both kinds of facial representations have their advantages and disadvantages. On one hand, feature-based techniques are more robust to variations in scale, size, head orientation, and location of the facial features, but usually they are computationally more expensive than the template-based techniques, and they require a very precise feature localization. In addition to these, psychological studies indicate that the human visual system processes faces at least to some extend holistically [3]. On the other hand, the template-based techniques have only very limited recognition and generalization capabilities, which may be caused by the smoothing of some important individual facial details, by small misalignment of the faces, and also by large interpersonal expression differences [3]. Therefore, many researchers use hybrid methods for expression recognition [1, 2].

Zhang et al. [6. 7] provided a hybrid method for expression recognition with a multilayer perceptron. They used two types of features. The first was the geometric position of 34 fiducial points. These points were selected manually and their coordinates form a feature vector of 68 elements. The second type of features was a set of Gabor wavelet coefficients at these 34 selected fiducial points. Each image was convolved with 18 Gabor filters (3 scales and 6 orientations), so there were in total 18 complex Gabor wavelet coefficients at each fiducial point and only magnitudes were used. Thus, the second feature vector contained 612 ($34 \times 3 \times 6$) elements. The two types of features were used both independently and jointly. Experimental results showed that Gabor wavelet coefficients were much more efficient than geometric position, and the highest recognition rates were obtained when both features were combined. Lyons et al. used a similar representation for the face and applied a wavelet of five spatial frequencies and six angular orientations. Then principal components analysis (PCA) was used to reduce the data, and finally they used a simple LDA-based classification scheme [8, 9]. Guo and Dyer also adopted a

[1] The text was submitted by the authors in English.

Received April 1, 2005

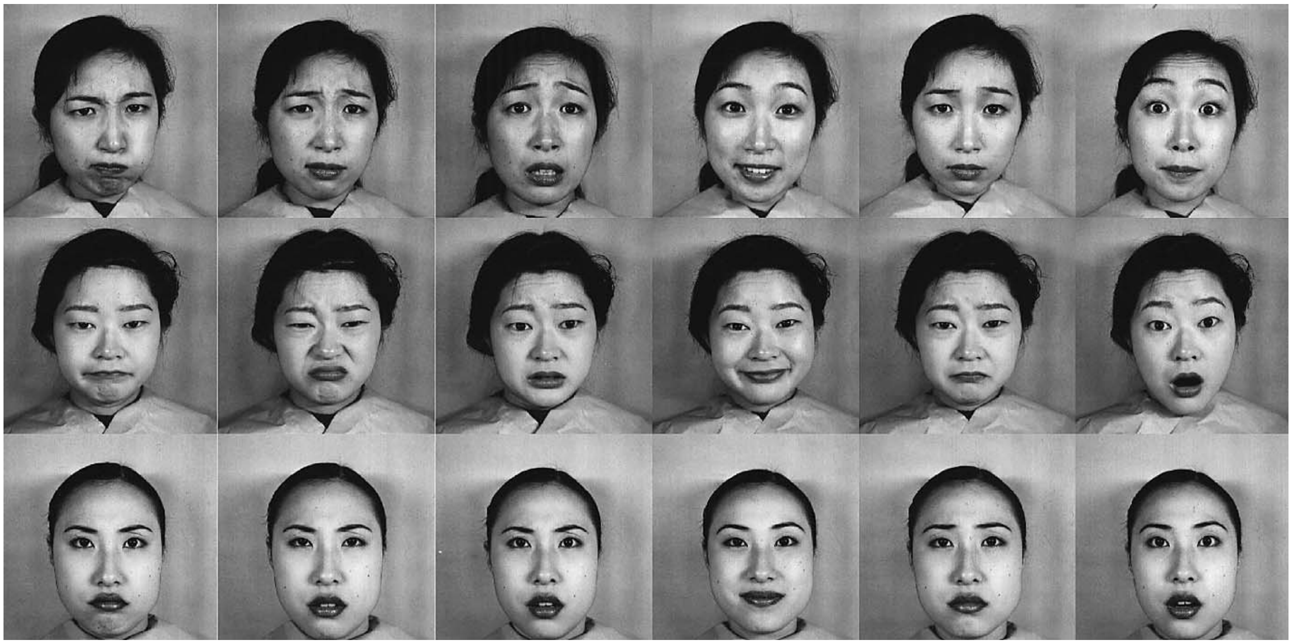

Fig. 1. Samples from the Japanese female facial expression image database. The resolution is 256×256 pixels.

similar face representation as Zhang et al., while they used a linear programming technique to carry out simultaneous feature selection and classifier training and got a better result [10]. All the three methods above required manual selection of 34 fiducial points. Buciu et al. used ICA and Gabor representation for facial expression recognition and also got a good result on the same database [4], but it is not clear how they divided the database. Extraction of facial features that can represent expressions effectively and also can be implemented easily is still one of the imperative questions for expression recognition.

Recently, a texture description based on local binary patterns (LBP) was developed. At first, very good performance was obtained in various texture classification and segmentation problems [11, 12]. Then, it was successfully extended to face recognition [13].

In this paper, we propose a novel hybrid method which can extract effective facial features easily. In the feature extraction step, the local binary pattern (LBP) operator is used to describe facial expressions. In the classification step, seven expressions are decomposed into 21 expression pairs such as anger–fear and happiness–sadness, so the seven-class classification problem is decomposed into 21 two-class classification problems. Twenty-one classifiers are produced based on the linear programming (LP) technique, each corresponding to one of the 21 expression pairs. A simple binary tree tournament scheme with pairwise comparisons is used here for classifying one kind of expression.

The rest of the paper is organized as follows. After briefly describing the image database in Section 2, we present our facial feature representation in Section 3. In Section 4, we introduce the expression classification method. In Section 5, experimental results are described. Finally, in Section 6, we conclude the paper.

2. FACIAL EXPRESSION DATABASE

The database we use in our study contains 213 images of Japanese female facial expression (JAFFE) [9]. Ten expressers pose three or four examples of each of the seven basic expressions (happiness, sadness, surprise, anger, disgust, fear, neutral). Sample images from the database are shown in Fig. 1.

In our study, image preprocessing is conducted by the preprocessing subsystem of the CSU Face Identification Evaluation System [14]: First, the images are registered using eye coordinates and cropped with an elliptical mask to exclude nonface area from the image. As a result, the size of each normalized image is 150×128. Then, the gray histogram over the nonmasked area is equalized and the pixel values are scaled to have a mean of zero and a standard deviation of one (see Fig. 2). Note that, owing to the invariance of the LBP operator to monotonic gray level changes, the gray scale normalization steps would not be necessary.

3. FACE REPRESENTATION WITH LOCAL BINARY PATTERNS

Emotion is more often communicated by facial movement, which will change visible appearance such as shape and position of facial features (eyebrows, eyes, mouth, etc.) and will also change textural structure of facial features and other regions in a face. With that knowledge it would be possible to analyze any facial

Fig. 2. Samples of the preprocessed images. The resolution is 150×128 pixels.

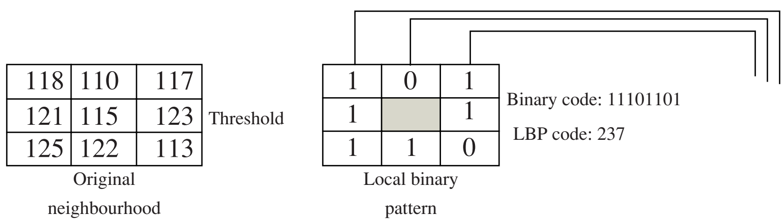

Fig. 3. The basic LBP operator.

movement by changes in facial appearances and then analyze facial expression. Our feature extraction method is just based on this point. In the following part, after introducing the LBP operator, we present our feature extraction method.

Figure 3 is an illustration of the basic LBP operator [12]. The original 3×3 neighborhood at the left is thresholded by the value of the center pixel, and a binary pattern code is produced. The LBP code of the center pixel in the neighborhood is obtained by converting the binary code into a decimal code.

On the basis of this operator, each pixel of an image is labeled with an LBP code by thresholding its neighborhood with the value of the center pixel [12]. The 256-bin histogram of the labels (distribution of pattern codes) contains the density of each label over a local region and can be used as a texture descriptor of the region.

The LBP operator introduced by Ojala et al. has been shown to be a powerful measure of image texture. It has been applied to many problems with excellent performance [11–13]. In our work, face images are seen as a composition of micropatterns, which can be well described by LBP.

We find that not all 256 patterns are equally important in our expression recognition. Some patterns have such low occurrence frequency that they contribute little to facial expression. The recognition results change little if we abandon these patterns. We observe that, for each image in our database, the "uniform" patterns (definition of "uniform" is in [12]) are the vast majority of patterns (over 96%) and we choose only the 58 uniform patterns of the 256 original patterns for facial representation.

Now, feature extraction is implemented with the following steps:

(1) Divide the face image into small regions. The size of each preprocessed image is 150×128. After experimenting with different block sizes, we choose to divide each preprocessed image into 80 (10×8) non-overlapping blocks (regions) of 15×16 pixels (see Fig. 4).

(2) Calculate the LBP histogram of each region. The LBP histogram of each region is obtained by scanning it with the LBP operator. Since only the uniform patterns are considered, we get a 58-bin histogram for each region.

(3) Concatenate each region's LBP feature histograms into a single feature vector. Now each region's 58-bin histogram is combined together to form a single feature vector for the whole image. Since 80 regions are used here, the feature vector has 4640 (58×80) elements.

The idea behind using our approach for feature extraction is motivated by the fact that, when facial features move toward some directions, both their position and shape will change. Accordingly, they or part of them will appear in different regions, resulting in some changes in LBP histograms in related regions. For example, when the eyebrows move upwards, they or part of them will leave their original regions and go into upper regions, so the related regions' LBP histogram will change. Because of the relationship between regions, other regions' LBP histogram will also change to some extent. So the LBP histogram is a good indicator of appearance changes because it can combine shape, position, and texture information.

The feature vector is still quite long and it is observed that some uniform patterns have very low occurrence frequency, so the feature vector needs to be reduced for speeding up the classification based on linear programming techniques.

In [10], a feature selection method is introduced which carries out feature selection and classifier training simultaneously. It is not used in our feature selection procedure because it is not suitable for very long feature vectors and it will lead to a degraded classification in theory. Details of this method are in Section 4.

A simple method is used here to erase these inessential patterns, which is based on the training samples:

(1) All of the training samples' feature vectors are summed up and then averaged by the number of training samples.

(2) For each region of an image, those patterns are discarded whose occurrence frequency is lower than a threshold, which is related to the size of the region. For example, set a threshold of 2.4 (since each region contains $15 \times 16 = 240$ pixels, the threshold means that those patterns whose average occurrence frequency is below (or equal to) one percent of the total number of all patterns in one region will be discarded).

After this procedure, the length of the feature vector is reduced greatly (to 1275 on average).

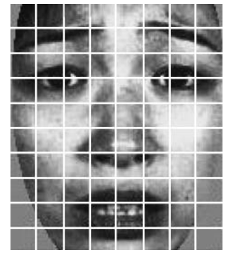

Fig. 4. An example of a facial image divided into 10×8 blocks.

4. EXPRESSION CLASSIFICATION BASED ON LINEAR PROGRAMMING TECHNIQUE

In [15, 16], a single LP formulation is proposed which generates a plane that minimizes an average sum of misclassified points belonging to two disjoint sets of points. We briefly describe this LP formulation below.

Consider two sets of points A and B in an n-dimensional real space R^n represented by the $m \times n$ matrix A and the $k \times n$ matrix B, respectively. The separating plane is as follows:

$$P := \{x | x \in R^n, x^T\omega = \gamma\}. \tag{1}$$

Here, $\omega \in R^n$ is normal to the separating plane with a distance $\frac{|\gamma|}{\|\omega\|}$ to the origin.

The separating plane P determines two open half-spaces $\{x|x \in R^n, x^T\omega > \gamma\}$ containing mostly points belonging to A and $\{x|x \in R^n, x^T\omega < \gamma\}$ containing mostly points belonging to B.

That is, we wish to satisfy

$$A\omega > e\gamma, \quad B\omega < e\gamma. \tag{2}$$

Here e is a vector of all 1s with appropriate dimension. To the extent possible, or upon normalization,

$$A\omega \geq e\gamma + e, \quad B\omega \leq e\gamma - e. \tag{3}$$

Condition (2) or (3) can be satisfied if and only if A and B do not intersect, which in general is not the case. We thus attempt to satisfy (3) by minimizing some norm of the average violations of (3) such as

$$\min_{\omega, \gamma} \frac{1}{m}\|(-A\omega + e\gamma + e)_+\|_1 + \frac{1}{k}\|(B\omega - e\gamma + e)_+\|_1. \tag{4}$$

Here, x_+ denotes the vector in R^n satisfying $(x_+)_i := \max\{x_i, 0\}$, $i = 1, 2, \dots n$. The norm $\|\cdot\|_p$ denotes the p norm, $1 \leq p \leq \infty$.

Formulation (4) is equivalent to the following robust linear programming formulation

$$\min_{\omega, \gamma, y, z} \frac{e^T y}{m} + \frac{e^T z}{k}$$

subject to

$$\begin{aligned} -A\omega + e\gamma + e &\leq y, \\ B\omega - e\gamma + e &\leq z, \\ y \geq 0, \ z &\geq 0. \end{aligned} \tag{5}$$

Recently, the LP, framework has been extended to cope with the feature selection problem [16]. Guo et al. also extend it to simultaneous feature selection and classifier training [10], called feature selection via linear programming (FSLP):

$$\min_{\omega, \gamma, y, z} \left(\frac{e^T y}{m} + \frac{e^T z}{k}\right) + \mu e's$$

subject to

$$\begin{aligned} -A\omega + e\gamma + e &\leq y, \\ B\omega - e\gamma + e &\leq z, \\ -s \leq \omega &\leq s, \\ y \geq 0, \ z &\geq 0. \end{aligned} \tag{6}$$

Classification performance of different methods (213 images with seven expressions). Here, Mean, Max, Min, and σ are the mean recognition rate, the lowest recognition rate, the highest recognition rate, and the standard deviation of recognition rate for 20 trials, respectively

Methods	Mean	Min	Max	σ
Zhang et al. [6]	90.1%	–	–	–
Guo et al. [10]	91%	–	–	–
Our method	93.8%	92.0%	96.3%	1.35%

The last term $\mu e's$ in (6) is to minimize $\|\omega\|_1$, which will cause a large number of 0 components in ω. If an element of ω is 0, the corresponding feature is removed. Thus, only the feature corresponding to nonzero components in the normal ω are selected after linear programming optimization.

In Guo's study, an average of 16.0 to 19.1 features were selected from the 612 original features.

There are two shortcomings for this method. First, it is not suitable for long feature vectors. For example, in our experiments, more than 50 feature vectors are used for training each of the 21 classifiers, while each feature vector has 4640 elements. According to formulation (6), a matrix with more than $(2 \times 4640 + 50) \times (2 \times 4640 + 50 + 1)$ elements is formed for producing a classifier, which is too large for computation. Second, it is easy to see from formulation (6) that the method is a trade-off between minimizing the average sum of misclassified points and minimizing the number of features. The first and second terms in formulation (6), which are used for minimizing the average sum of misclassified points, have a limit of zero, while the last term, which is used for minimizing the number of features, has a limit bigger than zero. A combination of these terms by simply adding them together will lead to a degraded classification in theory.

In our work, instead of using formulation (6) for both feature selection and expression classification, we use the method in Section 3 for feature selection and retain adopting formulation (5) as a classifier to minimize wrong classifications.

Since formulation (5) is only used for separating two sets of points, the seven-expression classification problem is decomposed into 21 two-class classification problems. In the training stage, 21 classifiers corresponding to 21 expression pairs are formed with 21 pairs of $\{\omega, \gamma\}$. In the testing stage, the feature vector of a testing sample is imported into these classifiers for comparisons.

5. EXPERIMENTAL RESULTS

To compare our method with earlier works, we also use the cross-validation technique over the 213 images. More precisely, we use the following procedure to report our result.

(1) Divide the database randomly into ten roughly equal-sized sets.

(2) Train the classifiers using data from nine sets. The remaining set is used for testing.

Fig. 5. Examples of disagreement.

(3) The above process is repeated so that each of the ten roughly equal-sized sets is used once as the test set.

(4) Average the results over all ten cycles as the recognition rate of one trial.

We repeat the above procedure 20 times and get in total 20 average results for these trials, of which the highest rate is 96.3%, the lowest rate is 92.0%, the mean of the 20 trials is 93.8%, and the standard deviation is 1.35% (see table).

Now we compare the recognition performance of our method to other published methods that use the same database (see table). In [6], a multilayer perceptron was used with 90.1% recognition accuracy. In [10], a technique named feature selection via linear programming was used and they achieved an accuracy of 91%.

It was reported in [6] that the expressers found it most difficult to pose fear expressions accurately and a human has more difficulty in recognizing fear. They achieved a rate of 92.3% when all fear images were excluded. In our experiments, a highest rate of 95.6%, a lowest rate of 93.4%, and an average rate of 94.6% is obtained when all fear images are excluded.

In [8], a result of 92% using linear discriminant analysis (LDA) was reported, but it included only nine people's face images and, hence, only 193 of the 213 images were used. We will not compare our results to it and other reports [4, 5, 17] with the same database because it is not very clear how they divided the database.

It should be pointed out that, in all three methods in [6, 8, 10], 34 fiducial points have to be selected manually. In our method, we need only the position of the two pupils for face normalization and other procedures are completely automatic. It is easy to see that our method performs much better.

It should also be noted that, in the JAFFE database, some expressions were expressed inaccurately or labeled incorrectly. This also influences our recognition result. Figure 5 shows a few examples with the labeled expression and our recognition results.

6. CONCLUSIONS

How to represent face efficiently is one of the most important questions for face and facial expression recognition. The local binary pattern operator, which has shown excellent performance in texture classification and face recognition, is used here to describe a face efficiently for expression recognition. As we have discussed, it combines position, shape, and texture information of a face and facial features. Then, 21 classifiers are produced based on a linear programming technique and classification is implemented with a binary tree tournament scheme. Experimental results demonstrate that our method performs better than other methods on the JAFFE database.

REFERENCES

1. M. Pantic and Leon J. M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State and the Art," IEEE Trans. on Pattern Analysis and Machine Intelligence **22** (12), 1424–1445 (2000).
2. B. Fasel and J. Luettin, "Automatic Facial Expression Analysis: A Survey," Pattern Recognition **36** (1), 259-275 (2003).
3. W. Fellenz, J. Taylor, N. Tsapatsoulis, and S. Kollias, *Comparing Template-Based, Feature-Based and Supervised Classification of Facial Expression from Static Images* (Computational Intelligence and Applications, World Scientific and Engineering Society Press, 1999).
4. I. Buciu, C. Kotropoulos, and I. Pitas, "ICA and Gabor Representation for Facial Expression Recognition," in *Proceedings Int. Conf. on Image Processing* (2003), pp. 855–858.
5. M. Gargesha and P. Kuchi, *Facial Expression Recognition Using Artificial Neural Networks* (EEE 511—Artificial Neural Computation Systems, Spring, 2002).
6. Z. Zhang, "Feature-Based Facial Expression Recognition: Sensitivity Analysis and Experiment with a Multi-Layer Perceptron," Int. J. on Pattern Recognition and Artificial Intelligence **13** (6), 893–911 (1999).
7. Z. Zhang, M. Lyons, M, Schuster, and S. Akamatsu, "Comparison between Geometry-Based and Garbor-Wavelet-Based Facial Expression Recognition Using Multi-Layer Perceptron," in *Proceedings 3rd Int. Conf. Automatic Face and Gesture Recognition* (1998), pp. 454–459.
8. M. Lyons, J, Budynek, and S. Akamastu, "Automatic Classification of Single Facial Images," IEEE Trans. Pattern Analysis and Machine Intelligence **21** (12), 1357–1362 (1999).
9. M. Lyons, S. Akamastu, M. Kamachi, and J. Gyoba, "Coding Facial Expressions with Gabor Wavelets," in *Proceedings Third IEEE Conf, Face and Gesture Recognition* (Nara, Japan. Apr. 1998), pp. 200–205.
10. G. D. Guo and C. R. Dyer, "Simultaneous Feature Selection and Classifier Training via Linear Programming: A Case Study for Face Expression Recognition," in *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition* (June, 2003), Vol. 1, pp. 346–352.
11. M. Pietikäinen, T. Nurmela, T. Mäenpää, and M. Turtinen, "View-Based Recognition of Real-World Textures," Pattern Recognition **37** (2), 313-323 (2004).
12. T. Ojala. M. Pietikainen, and T. Mäenpää, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Trans, on Pattern Analysis and Machine Intelligence **24** (7), 971–987 (2002),
13. T. Ahonen, A. Hadid, and M. Pietikäinen, *Face Recognition with Local Binary Patterns* (Lecture Notes in Computer Science 3021, Springer, 2004), pp. 469–481.
14. D. Bolme, M. Teixeira, J. Beveridge, and B. Draper, "The CSU Face Identification Evaluation System User's guide: Its Purpose, Feature and Structure," in *3rd International Conf. on Computer Vision Systems* (2003), pp. 304–313.
15. K. P. Bennett and O. L, Mangasarian, "Robust Linear Programming Discrimination of Two Linearly Inseparable Sets," Optimization Methods and Software **1**, pp. 23–34 (1992).
16. P. S. Bradley and O. L. Mangasarian, "Feature Selection via Concave Minimization and Support Vector Machines," in *Proceedings 5th Int. Conf. Machine Learning* (1998), pp. 82–90.
17. B. Fasel, "Head-Pose Invariant Facial Expression Recognition Using Convolutional Neural Networks," in *Proceedings of the Fourth Int. IEEE Conf. on Multimodal Interfaces* (2002), pp. 529–534.

Xiaoyi Feng received her doctor of technology degree in electrical engineering from the Northwestern Polytechnic University, China, in 2001. She is currently an associate professor at the Northwestern Polytechnic University. Her research interests include computer vision, image processing, and pattern recognition. She has about 20 scientific publications.

Abdenour Hadid is finishing his doctoral research at the University of Oulu, Finland. He graduated and received his engineer diploma from the National Institute of Informatics (INI), Algeria, in 1997. His research interest includes computer vision and pattern recognition. He is currently focusing on face detection and recognition, color image analysis, and learning. He has authored about 20 papers in international conferences and one journal article. He has served as a reviewer to several international conferences and journals such as *IEEE Transactions on Pattern Analysis and Machine Intelligence, International Journal of Information Visualization*, and International Conference on Image Processing (ICIP). He is member of the Pattern Recognition Society of Finland.

Matti Pietikäinen received his doctor of technology degree in electrical engineering from the University of Oulu, Finland, in 1982. In 1981, he established the Machine Vision Group at the University of Oulu. The research results of his group have been widely exploited in industry. Currently, he is professor of information engineering, scientific director of Infotech Oulu research center, and leader of the Machine Vision Group at the University of Oulu. From 1980 to 1981 and from 1984 to 1985, he visited the Computer Vision Laboratory at the University of Maryland, USA. His research interests are in machine vision and image analysis. His current research focuses on texture analysis, face image analysis, learning in machine vision, and machine vision for sensing and understanding human actions. He has authored about 165 papers in international journals, books, and conference proceedings, and nearly 100 other publications or reports. He is associate editor of the journal *Pattern Recognition* and was associate editor of *IEEE Transactions on Pattern Analysis and Machine Intelligence* from 2000 to 2005, He was chairman of the Pattern Recognition Society of Finland from 1989 to 1992. Since 1989, he has served as a member of the governing board of the International Association for Pattern Recognition (IAPR) and became one of the founding fellows of the IAPR in 1994. He has also served on committees of several international conferences. He is a senior member of the IEEE and vice-chair of the IEEE Finland section.