Constructing a Fuzzy Grammar for Syntactic Face Detection

A.Z. Kouzani, F. He, and K. Sammut
School of Engineering, Flinders University of SA, GPO Box 2100, Adelaide, SA 5001, Australia
e-mail: egazk, egfh, egks@eng.flinders.edu.au

Abstract— **This paper presents a structural face detection system. The proposed system consists of three stages; preprocessing, face-components extraction, and final decision-making. In the first stage, image conversion, colour operation, image restoration, and image enhancement are carried out. Face components are extracted in the second stage. A face model is defined, and a fuzzy grammar composed of octal chain codes is used to represent each of the seven face components. The practical limitations of this representation are considered. Structural components are detected, and the possibility degree that the extracted component is a real face component is determined. In the last stage, a commonsense knowledge base is employed for final evaluation. The detected face components and their corresponding possibility degrees allow the human face knowledge base to locate faces in the image and generate a membership degree for that face within the face class. The experimental results obtained using this method are presented.**

I. Introduction

The location of human faces in an image is the first difficult step of an automatic face recognition process. Although human beings locate a face and its components - eyebrows, eyes, nose, and mouth - within an image in a seemingly easy fashion, face detection is a complicated task to implement in a computer system. Some of the factors which make this task difficult are the presence of noise, occlusion, illumination, facial hair, make-up, spectacles, variation in scale, and orientation. Among the existing automatic face recognition systems [2]-[4], [6]-[9], [12]-[14], [16], [17], [19]-[22], only a few have addressed the problem of locating human faces in a scene before performing the recognition process [16], [19], [20], [22]. Most of the automated recognition approaches use centred input face images taken under constrained conditions, similar to those used to capture the database images. The existing automatic face recognition systems, which tackle the face detection problem, offer no solutions to most of the above-mentioned obstacles.

The human face is a complex pattern which contains most of its meaningful information in its structure. This characteristic poses an interesting challenge for computational modelling of the human face. The structural description of a pattern is a characteristic of syntactic pattern recognition [5], [18], in which a large set of complex patterns are described in terms of small sets of simple sub-patterns of primitives. These primitives are defined using grammatical rules that are derived from formal language theory. In practical situations, most patterns encountered are noisy and distorted. Fuzzy language is a suitable tool for describing the ill-defined structural information [1], [15]. In this case, the power of a grammar is increased by introducing fuzziness either in the definition of primitives, or in relations between primitives, or in both of these. A fuzzy grammar produces a language that is a fuzzy set of strings with the membership value of each string denoting the degree of belonging of the string in that language. The grade of membership of an unknown pattern in a class, described by the grammar, is obtained using a compositional rule of inference.

In face detection, the presence of pattern distortion, measurement noise, occlusion, and variation in patterns will lead to an erroneous string being rejected by the grammar characterising its class. Therefore, these limitations must be considered in the construction of the fuzzy grammars. However, it is still likely that some face patterns would not be detected by the grammars. A decision maker is therefore required to analyse the recognised patterns and to decide whether or not a face exists within an image. The decision maker employs a commonsense knowledge base [10] with domain knowledge of the human face to improve the detection results. The type of information that a commonsense knowledge base can represent, and reason from, is beyond the capability of fuzzy grammars. There are several methods for implementing knowledge-based systems. Connectionist models are one such method which have a good potential to satisfy the computational constraints. A fuzzy neural network implementation of a commonsense knowledge base [10] is therefore employed in the system proposed in this paper.

In this paper, a structural face detection system is proposed. The theory of fuzzy grammars for syntactic pattern recognition is summarised in section II of this paper. Section III presents the proposed face detection system, comprising the image preprocessing stage, the face-components extraction stage, and the final decision-making stage. The preliminary experimental results are discussed in Section IV, which is then followed by the concluding remarks.

0-7803-3280-6/96/$5.00

II. Fuzzy Grammars for Syntactic Pattern Recognition

In syntactic pattern recognition, a pattern is expressed by a sentence in a language which is specified by a grammar. A set of primitives is chosen to form the set of terminals of the grammar. The production rules of the grammar represent the relation between the primitives [5]. To handle vagueness in ill-defined patterns, fuzziness is introduced in the definition of primitives and production rules. A fuzzy grammar produces a language that is a fuzzy set of strings. Each string has a membership value denoting the degree of belonging the string in that language [15].

A. Fuzzy Language

An *alphabet* V_T is a finite set of symbols. A *string* x over V_T is a sequence of symbols $x = x_1, \ldots, x_n$. The *null string* Λ is the sequence with no symbols. The set of all strings, including Λ, over an alphabet is denoted by V_T^*. A *fuzzy language* (FL) with alphabet V_T is a fuzzy subset of V_T^* defined as $\sum_{x \in V_T^*} \mu_{FL}(x)/x$ where $\mu_{FL}(x)$ is the grade of membership of the string x in FL.

B. Fuzzy Grammars

A fuzzy grammar (FG) is a 6-tuple (V_N, V_T, P, S, J, μ) where V_N is a set of non-terminals, V_T is a set of terminals, P is a set of production rules, S is a starting point, $J = \{r_i \mid i = 1, \ldots, n,\ n = \text{cardinality of } P\}$ is the set of labels for production rules, and μ is a mapping $\mu : J \rightarrow [0, 1]$.

FG generates a fuzzy language (L(FG)) as follows. A string $x \in V_T^*$ is in L(FG) iff it is derivable from S, and its grade of membership $\mu_{\mathrm{L(FG)}}(x) = max_{1 \leq k \leq m}[min_{1 \leq i \leq l_k} \mu(r_i^k)]$ in L(FG) is > 0, where m is the number of derivations that x has in FG; l_k is the length of the kth derivation chain, and r_i^k is the label of the ith production used in the kth derivation chain, $i = 1, \ldots, l_k$. If a production $\alpha \rightarrow \beta$ is visualised as a chain link of strength $\mu(r)$, where r is the label of $\alpha \rightarrow \beta$, then the strength of a derivation chain is the strength of its weakest link, and therefore $\mu_{\mathrm{L(FG)}}(x)$ = strength of the strongest derivation chain from S to x for all $x \in V_T^*$.

C. Inference

In most applications of fuzzy grammars, the production rules and their membership degrees are predefined. However, an automatic mechanism can be employed to infer the fuzzy grammar from a specified fuzzy language. A reinforcement learning algorithm, the formal power series approach, and genetic algorithms are among the methods which can be used for this purpose.

D. Recognition

Given a fuzzy grammar FG describing the patterns of interest and a symbolic representation of x of an unknown pattern, *parsing* is necessary to recognise a fuzzy language FL. Parsing means deciding if $x \in L(FG)$. The recognition procedure fetches all production rules and executes all possible parsing for the input pattern. A pattern is recognised if its membership degree in the FL is greater than a threshold α.

III. The Proposed Face Detection System

The human face is a complex and meaningful pattern that contains most of its information in its structure. A human face is therefore expressed as a composition of its components. Although no two faces are alike, the front-view profiles of all human faces are similar. A model of the human face is thus defined in terms of its components and their relative positions as follows:

- A face consists of two eyebrows, two eyes, one nose, and one mouth.
- One eye is positioned left of the other eye.
- One eyebrow is above each eye.
- The nose is between and below the eyes.
- The mouth is below the nose.
- The distance between two eyes is approximately equal to the width of an eye.
- The width of the nose is smaller than the width of two eyes.
- The length of the nose is smaller than the width of three eyes.
- The width of mouth is smaller than the width of three eyes.
- The distance between the top of the mouth and the bottom of nose is smaller than the width of an eye.

Figure 1 illustrates the human face model. This model is utilised for structural face detection in this paper.

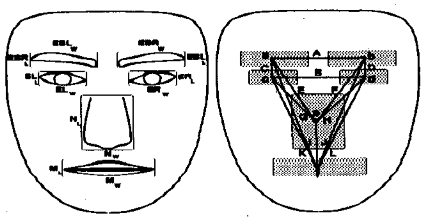

Fig. 1. Model of the human face.

The proposed face detection system consists of three stages; preprocessing, face-components extraction, and final decision-making. A block diagram of the system is shown in Figure 2.

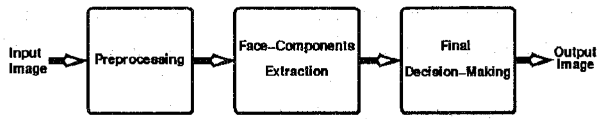

Fig. 2. The system block diagram.

A. Preprocessing

The functions of the preprocessing stage include image conversion, colour operation, image restoration, and image enhancement. The output of this stage contains an edge-extracted image of same sizes as the input image. Figure 3 illustrates a sample input image and the output of the preprocessing stage.

Fig. 3. A sample input image and the output of the preprocessing stage.

B. Face-Components Extraction

In this stage, the face components are extracted from the output of the preprocessing stage. The face components are eyebrows, eyes, nose, mouth, and face edges.

B.1 Primitive Extraction

Octal chain codes are used to describe the detected edges of the image. A unit length edge is represented by one of the directional octal code shown in Figure 4. The image is scanned first to find the pixels which represent the edges.

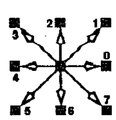

Fig. 4. Octal chain codes.

A string of octal code is then produced for each detected edge. The extracted string is next smoothened using a two-step process to remove undesired symbols caused by noise.

(i) The adjacent inverse code are deleted.
(ii) Within a group of three or four codes having zero total vector rotation, each pair is replaced by a digit or a pair of digits based on its combination.
Figure 5 illustrates some examples of the smoothing operation.

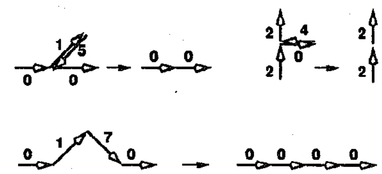

Fig. 5. Examples of the smoothing operation.

Each component of the face is described by a string of octal codes. The extracted primitives of the left-eye is explained below.

1. Iris:
$0^5\ 7\ 0\ (0\ 6)^5\ 7\ 6\ 6\ 7\ 6^5\ 5^3\ 4\ 6\ (6\ 4)^3\ 5\ 4\ 5\ 4^8\ (3\ 4)^3\ 3\ 2\ 4\ 2\ 4\ 2\ 2\ 4\ 2^7\ 1\ 2\ 2\ 1^5\ 0\ 1\ 0^4$.
2. Pupil:
$0^5\ 7\ 0^3\ 7\ 0\ 0\ 7\ 0\ (0\ 7)^{12}\ 0\ 6\ 6\ 5\ 4\ 4^3\ (5\ 4\ 4)^3\ 5\ 4\ 5\ 4\ 4\ 5\ 4\ 4\ 5\ 4^3\ 5\ 4^{10}\ 3\ 4^4\ (3\ 4\ 4)^7\ (3\ 4)^3\ (2\ 4)^3\ 2^4\ 1\ 1\ (0\ 1)^3\ 001\ 0^4\ 1\ 0^6\ 1\ 0^{11}$.

B.2 Construction of Fuzzy Grammars

In this part, context-free fuzzy grammars are constructed to be used in the recognition process. For each face component a specific fuzzy grammar is built. Both production rules and their membership grades are inferred from the fuzzy language which expresses the related face component. Therefore, the fuzzy languages must be determined first. The primitives of each face component extracted from the face model are used to determine the related fuzzy language. However, there are some practical limitations which must be considered in the construction of a fuzzy grammar for this particular application.

In face detection, the presence of pattern distortion and measurment noise will lead to an erroneous string being rejected by the grammar characterising its class. More importantly, the pattern which describes a face component varies from face to face and does not exactly follow the model described earlier. To build a fuzzy language for a face component, the related string for all allowable variations of the component model is first produced. The membership degrees of thses strings are set to 1. Then, the related strings of the distorted patterns are found from each string extracted in the previous step. The membership degrees of these strings are calculated from the distance between the child string and its parent. The child strings whose membership functions are below a threshold are ommited. Genetic programming [11] is employed for extraction of fuzzy languages. From the results it is found that the language generated for the left eye and the right eye are simillar. Therefore, only one language which is called eye is used for detection of both eyes.

Once fuzzy languages are determined, the related fuzzy grammars can be constructed. Genetic programming infers both production rules and their membership grades from the related fuzzy language. A minimal set of rules for generating the language is selected. The five inferred fuzzy grammars are left eyebrow, right eyebrow, eye, nose, and mouth.

B.3 Recognition

The recognition process fetches all production rules and executes all possible parsing for the input pattern. For each generated string a membership function is calculated using a max-min rule as stated in Section II. There exists a problem that must be considered in here. When an input image is preprocessed, the edges of a face component may not be extracted completely. Moreover, when there exists occlusion, an incompleted face component is extracted from the input image. As a result, a pattern will be rejected by the grammar. In our approach, this problem is solved in the recognition stage. When a sentence representing a reference pattern is matched against the input pattern, the operation is not terminated due to detection of the end of the input pattern. The process is however continued on the output of the preprocessing stage untill the end of the reference pattern is reached. If the input pattern is split into two or more parts, all parts will contribute to the calculation of the similarity measure. However, the missing parts can not contribute in the calculation of the similary criterion. The similarity criterion is the distance between the refernce string and the input string. If a measured similarity, which varies in the interval [0, 1], is greater than a threshold, the input string is said to be in the related language with a membership grade which is the measured similarity.

C. *Final Decision-Making*

The output of the face-component detector is fed into a commonsense knowledge-based system where the domain knowledge is used to analyse the input information and make the final decision. A connectionist model of commonsense knowledge representation and reasoning [10] is employed for implementation of the final decision-making stage. Fuzzy neurons are used to form the structure of the connectionist model.

The information provided for the face model forms the system domain knowledge. It describes a human face in terms of its components and their relationships. The information is provided into the knowledge base in two parts: the membership grades, and the relative spatial distances of the extracted face components. The membership degrees are directed into the input of *Components* attribute where a neuron is allocated to each face component. If a component is not recognised by its grammar, the output of the related neuron, varying in the interval [0, 1], is set to zero.

The relative distances of face components are fed into the *Relations* attribute in which the spatial relations among the face components are examined. A second order S function is employed to map the relative distances, varying in the interval $[0, 5\alpha]$, into $[0, 1]$. An input neuron is set aside for each distance variable. Other information can be also used for better verification of detected faces. For example, the area of the detected components, the colour of each region, etc. The knowledge base is implemented using a fuzzy neural network. A block diagram of the network is displayed in Figure 6. There are two attribute blocks and one conditional block shown in the figure. The *Components* block represents the propositions such as *Face has a left eye is true*. The *Relations* block represents the propositions such as *The distance between the centers of left eye and right eye is approximately equal to 2α*. The *If* block represents the conditional proposition such as *If components and relations then face*. The outputs of the *Components* and *Relations* attributes are examined in this block. The outputs denote a membership degree in the interval [0, 1]. This value shows the degree that the detected face components represent a face. The backward reasoners and the details about each block are not illustrated here, but are fully explained in [10].

IV. Results

The system has been implemented and tested on a large set of images. The face database contains more than 150 images of different scenes collected from the World Wide Web. The input images benefit from different spatial and gray scale resolutions. Each scene contains varying numbers of objects, including human faces, taken under varying illumination and orientation. Some input images contain different levels of noise. As it was stated in the previous section, the system evaluates a membership degree in the interval [0, 1] for each detected face. This value denotes the degree with which the detected face belongs to the

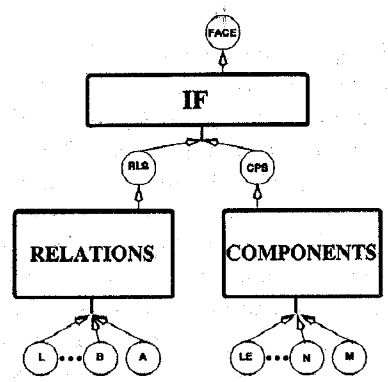

Fig. 6. The knowledge base block diagram.

set of human faces. Figure 7 displays the results obtained for a few image samples. Each detected face is surrounded by a rectangle. The thickness of a rectangle stands for the calculated membership degree of the detected face. Based on the experimental results, the proposed system could detect 83% of the faces in the face database. The experimental results show that the implementation of the face-component extraction stage should be reconsidered. More powerful preprocessing algorithms need to be implemented in order to overcome problems due to occlusion, noise, and illumination.

V. Conclusions

A system has been proposed in this paper for facial detection. An input image is preprocessed first for extraction of its edges. Then a syntactic approach is applied to the preprocessed image. It contains seven fuzzy grammars, each responsible for the recognition of a particular face component. The grammars are constructed in such a way as to minimise the effects of occlusion, noise, and illumination. A fuzzy neural network based on commonsense knowledge analyses the extracted face components, their membership degrees, and the domain knowledge that the system possess about human faces, and makes the final decision. Together with each detected face, a value is produced to denote the degree of membership of the face within the face class. The system is assessed based on a face database of more than 150 images of different scenes taken under varying conditions. The preliminary experimental results are promising, however there is a room for improvement of the system.

Acknowledgments

The first author would like to thank DEET, Australia, and FUSA, for providing him with scholarships to undertake this research.

References

[1] J.C. Bezdek and S.K. Pal, *Fuzzy Models for Pattern Recognition.* IEEE Press, 1992.

[2] D.J. Beymer, "Face recognition under varying pose," *TR-AI 1461*, Massachusetts Institute of Technology, December, 1993.

[3] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705-740, May, 1995.

[4] I.J. Cox, J. Ghosn, and P.N. Yianilos, "Feature-based face recognition using mixture-distance," *TR-95-09*, NEC Research Institute, October, 1995.

[5] K.S. Fu, *Syntactic Pattern Recognition, Applications.* Berlin: Springer-Verlag, 1977.

[6] G. Gordon, "Face recognition based on depth maps and surface curvature," *Geometric Methods in Computer Vision, Proc. SPIE*, 1570, 1991.

[7] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103-108, January, 1990.

[8] T. Kanade, *Picture Processing System by Computer Complex and Recognition of Human Faces.* PhD Thesis, Dept. of Information Science, Kyoto University, November, 1973.

[9] T. Kohonen, *Self-Organization and Associative Memory.* Berlin: Springer-Verlag, 1984.

[10] A.Z. Kouzani, F. He, and K. Sammut, "Commonsense knowledge representation and reasoning with fuzzy neuarl networks," *Submitted to the Fourth Australian and New Zealand Conference on Intelligent Information Systems (ANZIIS-96)*, Adelaide, Australia, 18-20 November, 1996.

[11] J.R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* Cambridge, MA: MIT Press, 1992.

[12] P. Kruizinga and N. Petkov, "Optical flow applied to person identification," *Proc. of the 1994 EUROSIM Conference on Massively Parallel Processing Applications and Development*, Delft, The Netherlands, pp. 871-878, June, 1994.

[13] O. Nakamura, S. Mathur, and T. Minami, "Identification of human faces based on isodensity maps," *Pattern Recognition*, vol. 24, no. 3, pp. 263-272, 1991.

[14] A.J. O'Toole, H. Abdi, K.A. Deffenbacher, and D. Valentin, "Low-dimensional representation of faces in higher dimensions of the face space," *Optical Society of America*, vol. 10, no. 3, pp. 405-411, March, 1993.

[15] S.K. Pal, D. Majumder, and K. Dwijesh, *Fuzzy Mathematical Approach to Pattern Recognition.* New Delhi: Wiley Eastern Limited, 1986.

[16] H.A. Rowley, S. Baluja, and T. Kanade, "Human face detection in visual scenes," *CMU-CS-95-158R*, School of Computer Science, Carnegie Mellon University, November, 1995.

[17] A. Samal and P.A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: A survey," *Pattern Recognition*, vol. 25, no. 1, pp. 65-77, 1992.

[18] M. Sonaka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision.* Chapman & Hall, 1993.

[19] K.K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *TR-AI 1521*, Massachusetts Institute of Technology, December, 1994.

[20] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.

[21] D. Valentin, H. Abdi, A.J. O'Toole, and G.W. Cottrell, "Connectionist models of face processing: A survey," *Pattern Recognition*, vol. 27, pp. 1209-1230, 1994.

[22] A.L. Yuille, "Deforable templates for face recognition," *Journal of Cognitive Neurosience*, vol. 3, no. 1, pp.59-70, 1991.

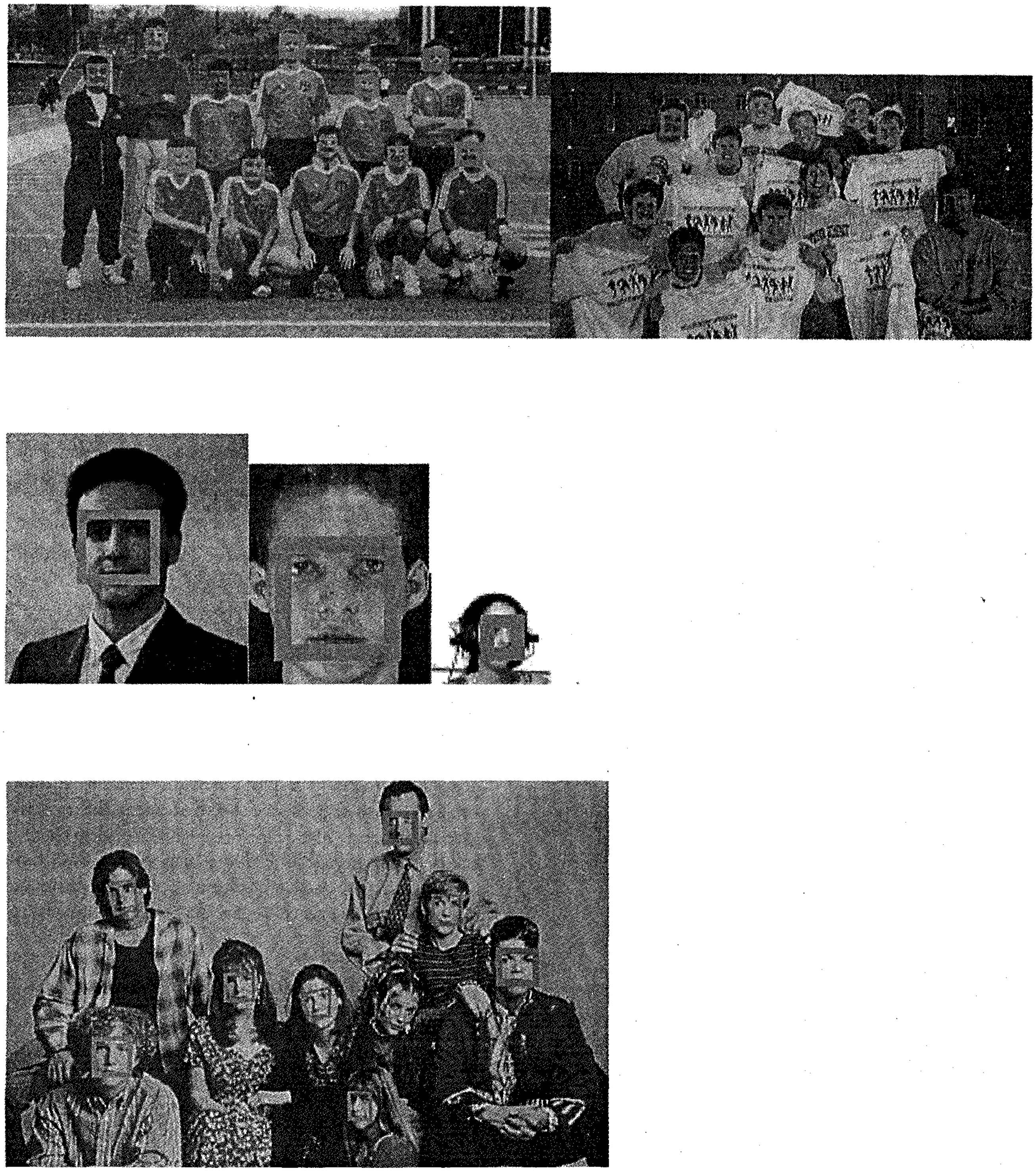

Fig. 7. Samples of the experimental results.