Pergamon

Pattern Recognition, Vol. 27, No. 1, pp. 53–63, 1994
Elsevier Science Ltd

0031-3203/94 $6.00 + .00

HUMAN FACE DETECTION IN A COMPLEX BACKGROUND

GUANGZHENG YANG† and THOMAS S. HUANG‡
† Department of Automation, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China
‡ Beckman Institute, University of Illinois at Urbana–Champaign, 405 North Mathews Ave, Urbana, IL 61801, U.S.A.

(*Received* 18 *December* 1992; *in revised form* 15 *June* 1993; *received for publication* 6 *August* 1993)

Abstract—The human face is a complex pattern. Finding human faces automatically in a scene is a difficult yet significant problem. It is the first important step in a fully automatic human face recognition system. In this paper a new method to locate human faces in a complex background is proposed. This system utilizes a hierarchical knowledge-based method and consists of three levels. The higher two levels are based on mosaic images at different resolutions. In the lower level, an improved edge detection method is proposed. In this research the problem of scale is dealt with, so that the system can locate unknown human faces spanning a wide range of sizes in a complex black-and-white picture. Some experimental results are given.

Face detection Knowledge-based pattern recognition Mosaic image Edge detection

INTRODUCTION

Human faces represent one of the most common patterns in our vision. The automatic recognition of human faces is a significant problem in the development and application of pattern recognition. In recent years this problem has attracted considerable attention.[1–14]

A complete human face recognition system should accomplish the following tasks:

(1) For an arbitrary picture, determine whether it contains any faces. If so, determine the number of faces as well as their position and size.
(2) Identify a person from his/her face.
(3) Make a description of facial expression (smile, surprise and so on).
(4) Make a description of each face. Find a certain face according to a given description.

This is a quite complex and difficult problem. Most research work to date can be regarded as only a starting point. Many aspects of human face recognition remain to be solved.

So far, much of the computer vision research in the field of human face recognition has focused on the identification task.[1, 5,7–9,14] In references (2, 8) authors proposed methods to recognize human face profiles. Most authors use a picture with a simple background or an interactive method to determine the face location. Thus, these methods are greatly restricted in their application.

However, several interesting methods for human face recognition have been proposed, for example, the methods using neural networks,[1,5,13] the methods using mosaic images,[4,14] and the methods using eigenfaces.[3]

From the above-mentioned tasks of face recognition, the first important step of fully automatic human face recognition is to detect and find the positions of faces in a given unknown picture (automatic human face location). This is one of the key problems in face recognition, so we must pay attention to it. In fact, automatic human face location itself is a problem of pattern recognition. It includes the automatic segmentation (find the region of a face and eye, nose, and mouth regions) and recognition (i.e. to distinguish a human face from other kinds of patterns in the background). Since both the location and the size of a face in a picture are unknown, the problem is very difficult.

The task of automatic face location in a complex background has been relatively unexplored. Most researchers used images with a simple background, and others limit the number of faces to one. The issue of scale remains largely unaddressed. Other work used additional a priori information, for example, captions.[10–12]

In this paper, we attempt to develop an automatic face location system which can find the locations of human faces in an image when the number, the location, and the sizes of faces are unknown, so that the system can be used for images with a complex natural background.

The system presented in this paper is a hierarchical knowledge-based pattern recognition system. A general block diagram is shown in Fig. 1.

The face location system consists of three levels (steps). The original digital image is the only input information to the system. At level 1, according to a set of rules, the whole input image is scanned globally to find all the possible candidates of human face with all possible sizes and locations. The face candidates obtained in level 1 then are the input to level 2. At level 2, a second set of face detection rules is established on

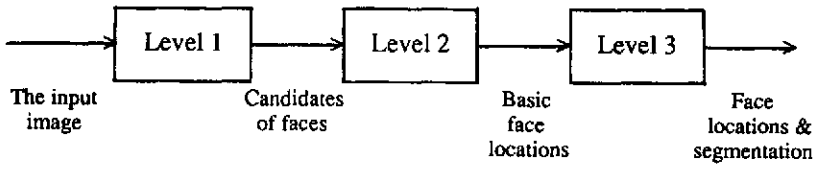


Fig. 1. General block diagram.

the basis of a window of 8 × 8 cells over each candidate face, and according to these rules the candidates are screened. The rules for levels 1 and 2 are established on the basis of mosaic images, just as in references (4, 14). At level 3 the candidates which survived from level 2 are screened again. An improved edge detection method for eye and mouth regions is proposed. If the extracted characteristics in level 3 fit well with the characteristics of eyes and mouth, the detection of a human face will be declared.

Our research is on the detection of human faces in black-and-white still images. Of course, if color still images or image sequences (video) are available, we can use also the additional color information or video characteristics to improve the face detection result. However, in many cases, the only available data are black-and-white still images; for example, in archiving existing images (in the context of content-addressable systems). Furthermore, the problem presents a great challenge: we humans have no trouble at all in locating faces in black-and-white still images; can we make a computer do it?

USE OF MOSAIC IMAGES

1. *Construction of mosaic images*

The rules of levels 1 and 2 are established on the basis of mosaic images. A mosaic image is constructed by decreasing the resolution of the original picture. A mosaic image consists of cells. For convenience of calculation, the original image is divided into square cells with the same size. In each cell there are $n \times n$ pixels, where n is the length of a side of a cell. The grey level of each cell equals the average value of grey levels of all $n \times n$ pixels included in this cell.

In face recognition two kinds of features are extracted: the first ones are line features, they show the outlines of a human face or organs in a face; the second are region features—the characteristics of the face region and eye and mouth regions are generally different from those of other regions in the image. As in face location the human faces have to be distinguished from their background, the region features are important for face location. The low-resolution characteristics of the mosaic images are suitable for extracting the region features, so in levels 1 and 2 we use the mosaic images to establish the rules for searching the human faces.

In Fig. 2 a series of mosaic images for an original image of 512 × 512 pixels is shown, while the lengths of the cell side are changed from $n = 1$ (the original image) to 128. The series of mosaic images shows that from the mosaic image with $n = 128$ we cannot get any conclusion whether there is a face in the picture or not. As the cell size becomes smaller, the macroscopic features of the human face will appear in a mosaic image. When $n = 64$ and 32, possible candidate positions of faces may be obtained. When $n = 16$, the human faces appear obviously. The face features are clearly visible at $n = 8$. When $n = 4$ or less, the facial expressions also can be seen.

Figure 2 shows only a special case. For another picture with a human face the obvious appearance of the face may happen not at $n = 16$, but at some other position of the series of the mosaic images (e.g. see Fig. 4). Moreover, if in a picture there exist several human faces of different sizes, they can appear obviously at different positions of the series of mosaic images respectively for this picture.

For the example shown in Fig. 2, if we take the mosaic image with $n = 32$ as the basis for searching in level 1, we can roughly find the possible candidates of a human face. Here in the mosaic image of $n = 32$, the main part of the face occupies an area of about 4 × 4 cells. According to this, we can construct a local mosaic image, which has 4 × 4 cells and occupies the main part of the face. This mosaic image is called a quartet. Analogously, for the example of Fig. 2 in the mosaic image of $n = 16$ the human face appears obviously. The mosaic image with $n = 16$ can be regarded as the basis of searching the human face in level 2. In this case, the human face occupies an area of about 8 × 8 cells. The mosaic image, which occupies the main part of the face and contains 8 × 8 cells, is called an octet. The quartet and the octet are useful images for face searching. On the basis of the quartet and octet the rules for levels 1 and 2 are established, respectively.

Figure 3 shows the quartet (Fig. 3(d)) and the octet (Fig. 3(c)) for the example of Fig. 2. Figure 3(a) is the original. Figure 3(b) is a mosaic image of Fig. 3(a) with side length $n = 32$. On Fig. 3(b) the searching in level 1 is processed by using the rules based on the quartet (Fig. 3(d)).

Figure 4 shows the quartet (Fig. 4(d)) and the octet (Fig. 4(c)) for another example with a face, the size of which is smaller than the one in Fig. 2. Figure 4(a) is the original. Figure 4(b) is a mosaic image of Fig. 4(a) with side length $n = 18$. From Fig. 4(c) we find that the face appears obviously when $n = 9$ (octet).

2. *Search process in the mosaic image*

If the rules established on the octet are complete and perfect, the octet of any face must be matched with

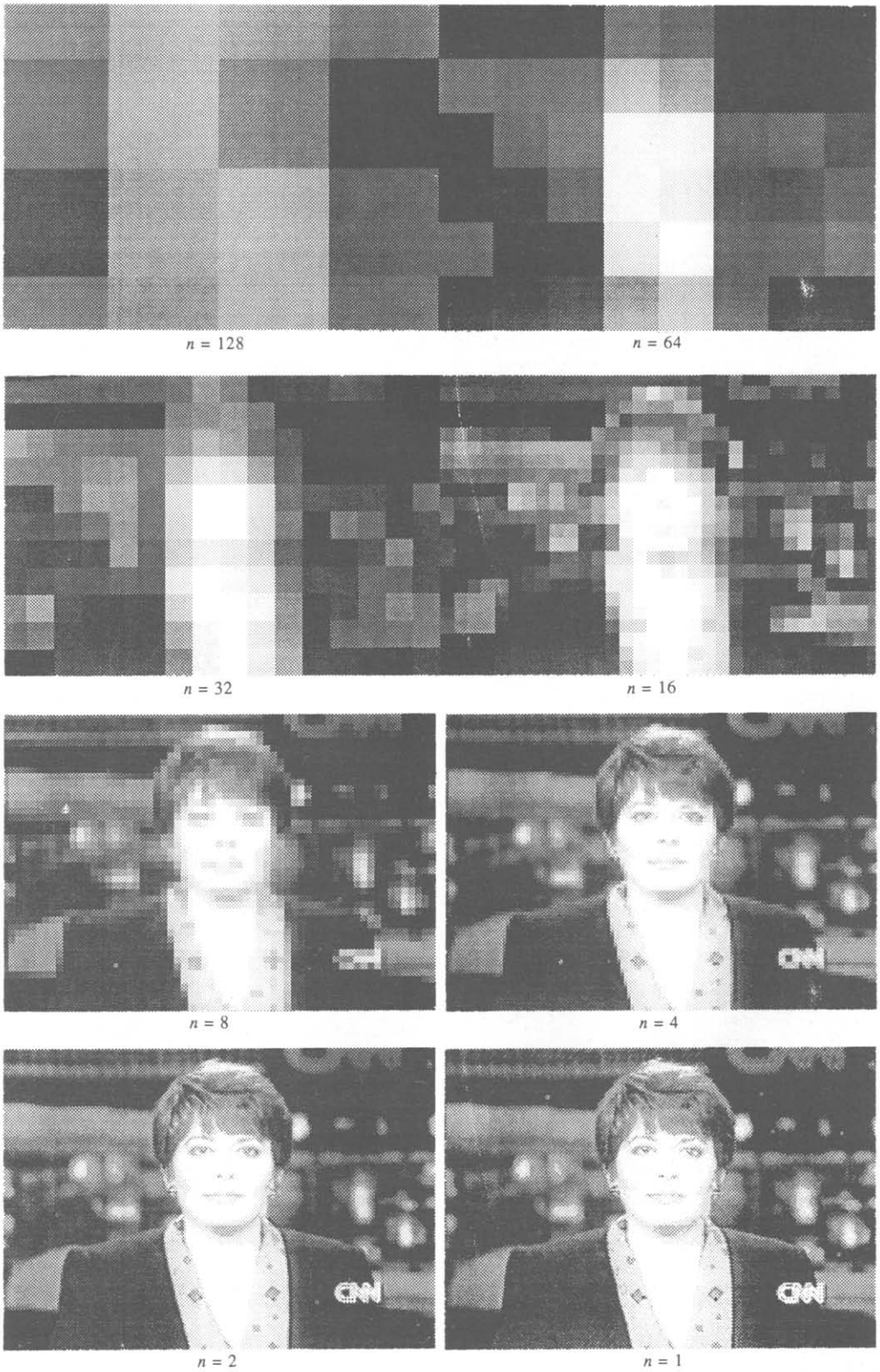


Fig. 2. Series of mosaic images (from $n = 1$ to 128).

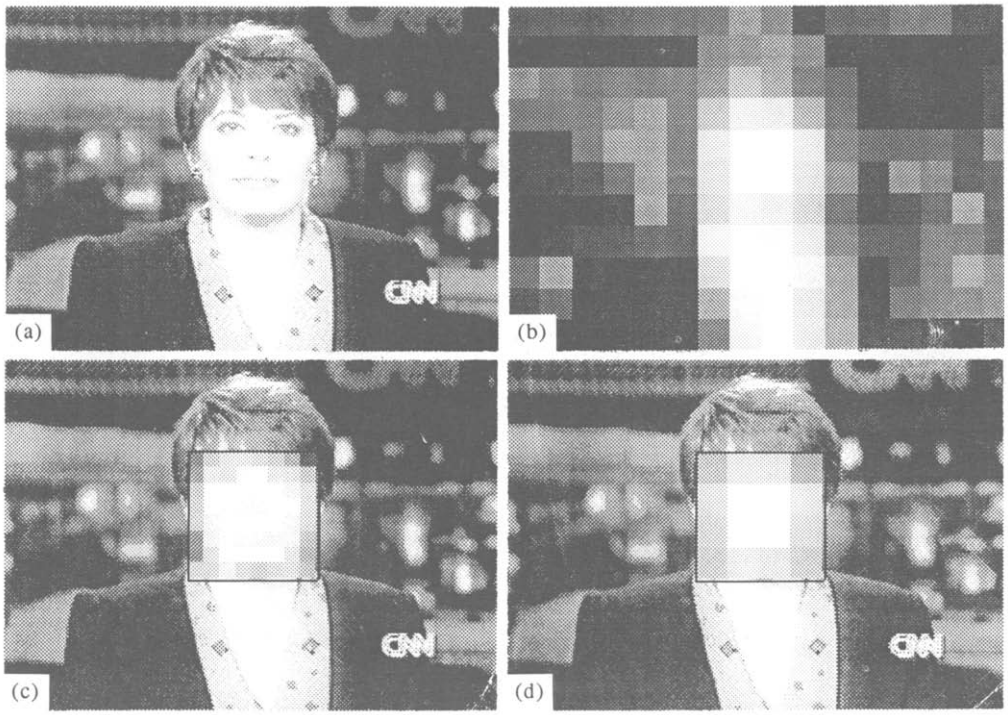


Fig. 3. The quartet and the octet for Fig. 2.

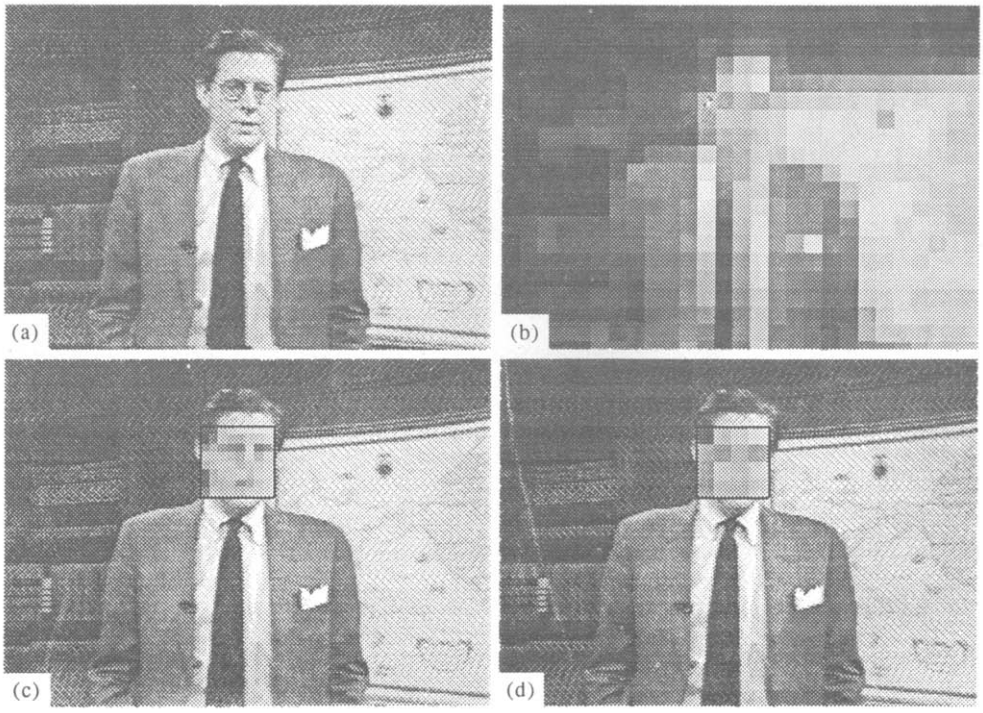


Fig. 4. The quartet and the octet for another picture.

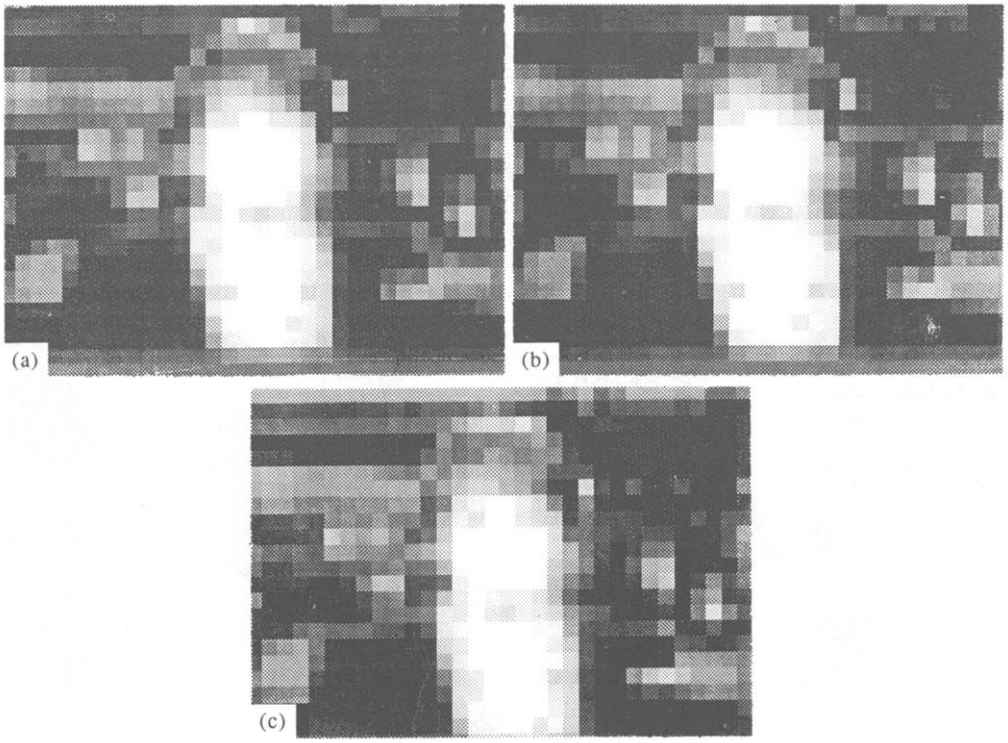


Fig. 5. Translation of the mosaic images.

these rules. Conversely, if a quartet and octet exactly fit the rules, there must be a face where the quartet and octet are located. Of course, such an ideal set of rules is difficult to obtain. Thus, we must tolerate the existence of errors.

Suppose the size of the face is known roughly, but the face location and the number of faces are unknown. We can construct a mosaic image with such a side length of cells that it just equals the length of the quartet cell according to the given face size. Then we search the face candidates on the constructed mosaic image using the rules of level 1. Next the octets of face candidates from level 1 will be used for the searching in level 2.

Now suppose that the size of the face is unknown. The system will in a certain range (in our experiment the range is from $n = 6$ to 25) construct a series of mosaic images, and on the series of mosaic images globally search the face candidates according to the rules of level 1. If in a location of the mosaic image with $n = k$, $6 \leq k \leq 25$, the rules of level 1 fit well, there will exist a quartet for a candidate with the length of cell side k. Thus, such a candidate is found in level 1. For all of these face candidates with length of cell side in the range from $n = 6$ to 25, level 2 searching is processed. In level 2 the corresponding octets of the candidates obtained from level 1 are constructed, and the existence of faces is tested by using the rules of level 2. Thus we have dealt with the problem of face location when both the size and location of faces in a picture are unknown, and we can get the information on the number, the position, and the sizes of faces in a picture.

The first two levels of our human face location system are suitable for searching faces with size of 8×8 mosaic cells. In fact, after digitization any picture can be regarded as a mosaic image with cell side length of one pixel. So all rules based on the quartet and octet are available for the case when $n = 1$. Thus, $n = 1$ is the lower limit of face size for levels 1 and 2. If $n = 1$, the corresponding face size just is 8×8 pixels.

As the locations of faces and other patterns in a picture are random, we have to take care of the effect of translation on mosaic images. The result of translation may change the mosaic image (Fig. 5), yet the change is small, and is periodic with a period equal to the side length of quartet cells. So it can be controlled. In the algorithm the effect of translation has been considered. Figure 5(a) shows the original mosaic image, Fig. 5(b) shows the mosaic image when the coordinates are moved to the right by 2 pixels, and Fig. 5(c) shows the case when the coordinates are moved down by 8 pixels.

EXTRACTION OF EDGES

The face searching in level 3 is based on extracting edges of organs in the human face. For the face candidates received from level 2 local histogram equalization is performed first. Thus, edges of the eyes and mouth may be detected using fixed thresholds.

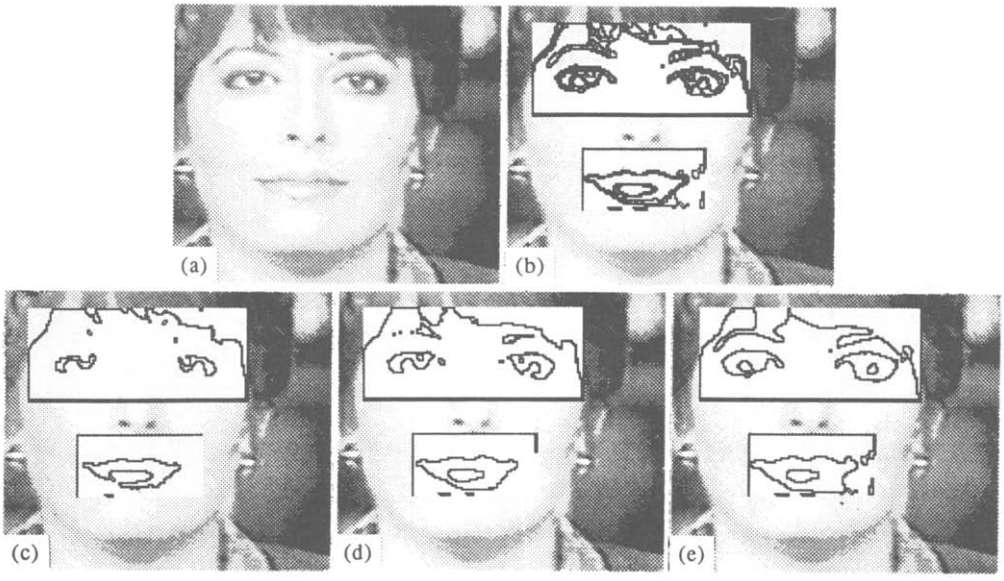


Fig. 6. Edge detection using multi-binarization.

A multi-binarization method is used to detect edges. Three different thresholds are applied, and results of the three edge sets are combined together. Figure 6 shows an example of edge detection using such multi-binarization. Figure 6(a) shows the original image, Figs 6(c)–(e) show the edges extracted using the thresholds 50, 75, and 100, respectively, Fig. 6(b) shows the result of combining all the edges from Figs 6(c) to (e).

For detecting a human face further feature extraction is needed. In this system the features for decision in level 3 are the length, slope, curvature, concavity, center of gravity, and the coordinates of the ends of an edge. For extracting these features an edge tracking algorithm is used. In level 3 only the eyes' and the mouth's edges are studied. As the eyes and the mouth are oriented mainly in the horizontal direction, the horizontal direction is emphasized in the tracking. Moreover, if a candidate from level 2 is a false face, prominent vertical edges may appear (e.g. the collar). For eliminating this type of false faces, an auxiliary vertical tracking is done to emphasize the features in the vertical direction.

DETECTION ERRORS

As the number of selected features is limited and the rules in levels 1 and 2 are not complete, we must tolerate errors. Just as for any recognition problems there are two kinds of errors in face detection:

(1) There exists a face in the picture, but a decision of "No face" is made.
(2) There is no face, but a decision of "Face" is made.

These two kinds of errors usually have different consequences. For the face location problem, if at the higher level one makes a decision of "Face" when there is no face, the false face may be eliminated later at a lower level by using new features. But if a face was lost at the higher level, the true face will never be found at the lower levels. So in our hierarchical face location system, we would rather have an error of the second kind at the higher level than the first kind.

DETAILS OF THE ALGORITHM

In levels 1 and 2 the search work is based on the quartet and octet in the mosaic images. The output of level 1 is a set of face candidates in the input picture. The candidates are obtained via global searching by using a set of simple rules based on the quartet. These rules reflect the principles mentioned above. The main idea of these rules is that the candidates are found by the contrast between the face and environment around the face. Examples of these rules are:

- The center part of the face (the quartet) has four cells with a basically uniform grey level (see Fig. 7 the darkly shaded part).
- The upper round part of a face (Fig. 7, the lightly shaded part; there are 8 cells altogether) has a basically uniform grey level.
- The difference between the average grey levels of the center part and the upper round part of the quartet is significant.

For level 2 a set of face detection rules is established on the basis of the octet. For example:

Eye rules. In the horizontal direction calculate the average value of the grey level of all eight cells for each row, and use this average value for the vertical searching. There may be an eye region, if

- there is a local minimum of derivative of the mean grey levels; or
- there is a local minimum of grey levels;

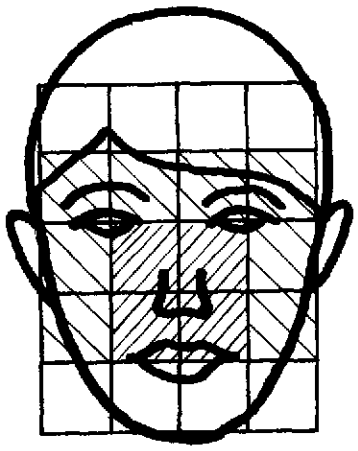


Fig. 7. A quartet for level 1.

and in the horizontal direction if

- there are two local minima of grey levels, the distance between these minima d: $2 < d < 5$; or
- there is one local minimum of grey levels, the brightest point is near the center and the distance between the minimum and the brightest point d: $0 < d < 3$.

Nose rules. There may be a nose region, if

- there is a local minimum of grey levels in the vertical direction under the eye centered at 1 unit of distance from the center, and the grey level is near the local minimum in the horizontal direction.

Mouth rules. There may be a mouth region, if

- there is a local minimum of grey levels in the vertical direction under the nose at 1 unit of distance, and in the horizontal direction the grey level has a 2–3 unit region with low grey levels.

For the same face, whether its size is large or small, its quartet and octet are the same. Furthermore, for a large class of images, the grey level variations in the quartets and octets of different faces can be characterized by a relatively small set of rules.

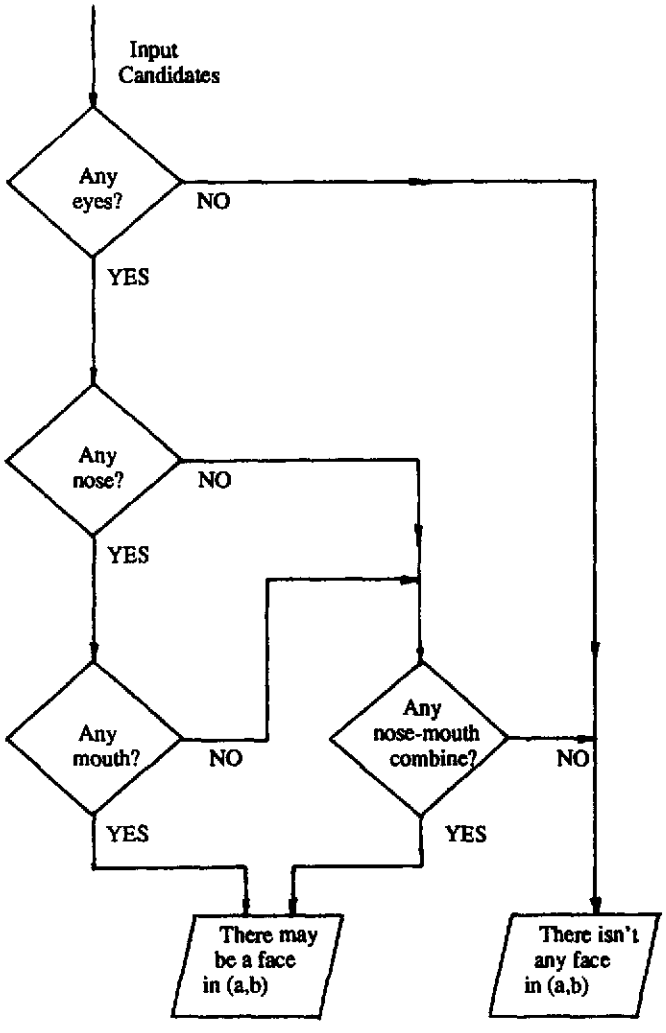
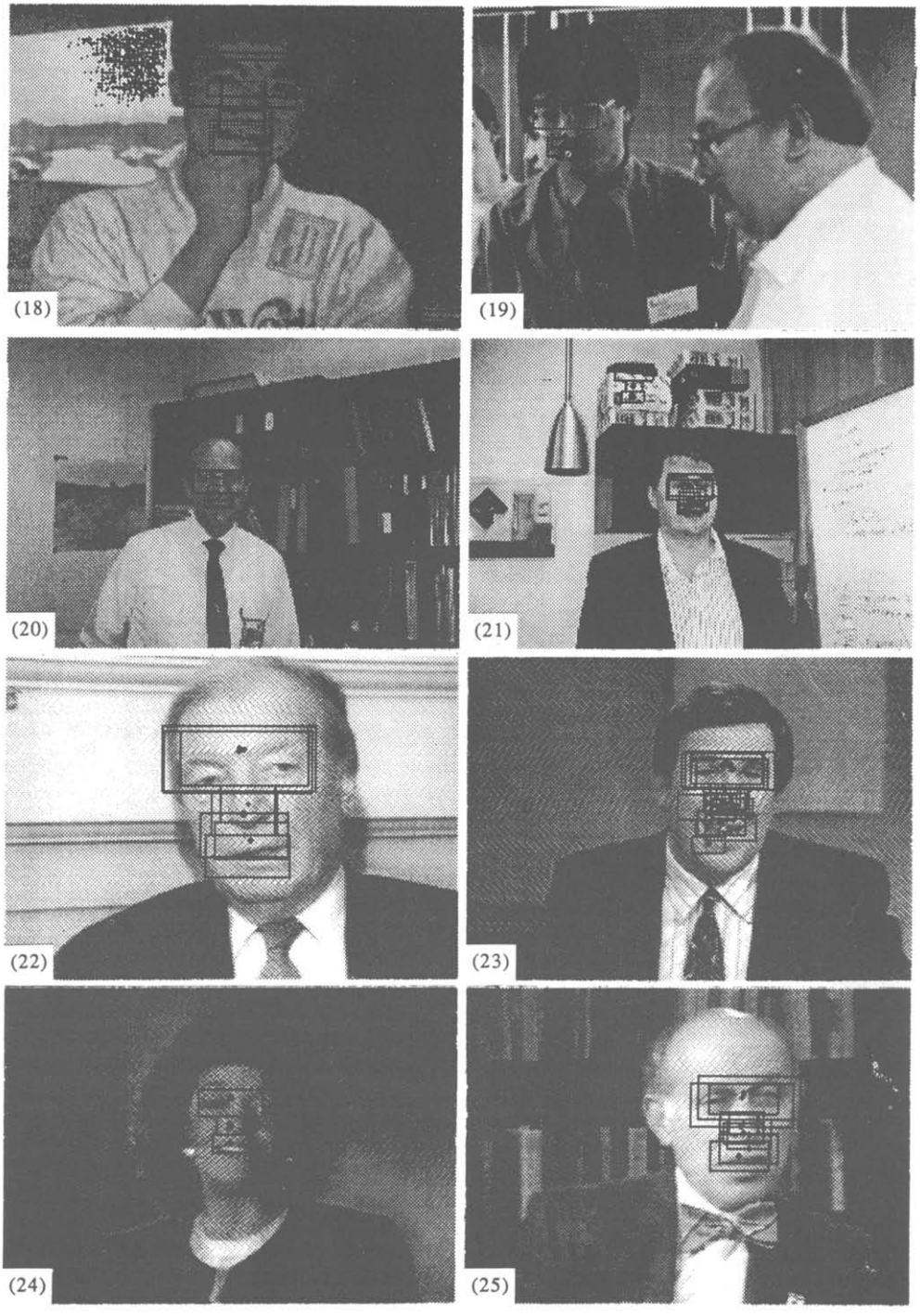


Fig. 8. Block diagram of level 2.

Fig. 9. Block diagram of level 3.

Figs 10–17. Examples of face location.



Figs 18–25. Examples of face location.

The rules based on the quartet and octet are obtained by comparing the relative grey levels. So the rules are quite simple and can be relatively independent on lighting conditions to a certain degree.

However, these rules are obtained by a training set consisting of only 40 pictures, and some cases of faces have not been considered (e.g. the tilted or turned faces, the faces with shades and so on), so the rules are not complete. We can improve this system by using more training examples and considering more face cases.

The results of level 2 may include false faces in addition to true ones. These results are then sent to level 3 for the final decision. A block diagram of level 2 is shown in Fig. 8.

In level 3 the candidates are screened again. A block diagram of level 3 is shown in Fig. 9. In level 3 local

histogram equalization, multi-binarization, and bi-directional edge tracking are used. Then the features of the edges of eyes and mouth are extracted. The main task of level 3 is to eliminate the false faces obtained from level 2. In our system after the edge detection and tracking, the first step is to deny all the obvious no faces. There are several simple rules for this goal, for example:

- if there are too many horizontal edges in the mouth region, it is never a mouth in this region;
- if there are too many vertical edges in the eye region, it is never eyes in this region.

Then if the extracted characteristics fit well with the characteristics of eyes and mouth, the conclusion of a human face will be made.

EXPERIMENTAL RESULTS

We used a set of 40 pictures as the training set and a set of 60 pictures as the test set. All rules and parameters used in our system are obtained from the training set. The pictures from the test set are used only for examination. With a few exceptions, most of our pictures contain only one face. Some of them were obtained from TV and the others were from photographs. From two pictures of the training set we created two subsets for translation. The subsets were obtained by translating the original to the right, down, and in the diagonal direction by steps of two pixels. So the two subsets contain 46 pictures each. The size scan is from $n = 6$ to 25. This is a quite large range for searching for a human face. For a picture with 512×512 pixels it means to allow for face sizes of 48×60 to 200×250 pixels.

The test set consists of 60 pictures. After level 3 the system successfully located 83% faces (50 from 60). There are 28 pictures in which besides the correctly located faces, false faces appeared.

As the candidates in level 1 are obtained according to the difference between the average grey levels of the center part of the face and the background, the system can detect both the dark face in a light background and the light face in a dark background, only if the difference of grey levels between the face and the background is more than a given threshold.

There are 10 pictures from the test set where the true faces are not found. One of them lost the true face on level 2. The other 9 pictures lost true faces on level 3. Five of them are eliminated by unmatching in the horizontal direction, and four of them by unmatching in the vertical direction. In these nine pictures, two of them are eliminated by unmatching in the eye's region, and seven of them are eliminated by unmatching in the mouth's region.

Since the task of automatic face location in a complex background has been relatively unexplored, we know of only a few works on face location. In reference (12) a 73% successful location rate for newspaper photographs was reported. Meanwhile, there are on average approximately two false alarms per picture. Another work is reference (13), where the neural networks were used. In this method in a picture the human faces are limited to almost the same size. The successful location rate was not reported clearly.

In the existing algorithm we did not consider explicitly the case of tilted or turned faces. We can put additional rules into the knowledge base to deal with these cases if necessary. But the existing algorithm as it is tolerates slight tilt and turn.

One of the most important reasons for errors is that in this algorithm we use only the relative value of a cell's grey level, and the rules in levels 1 and 2 are not complete, and in level 3, the feature extraction has not yet been perfected.

The run time for locating the face in a picture of 512×512 pixels is about 60–120 s on a SUN-4 Sparc station 2.

Figures 10–25 show the results of face location. Usually a face is detected at several different (but close) scales, because the rules may be matched for several scan sizes.

CONCLUSIONS

(1) Our algorithm is an efficient method of finding face locations in a complex background when the size of the face is unknown. It can be used for black-and-white pictures with a wide range of face sizes and does not need a priori information.

(2) We use a knowledge-based recognition approach. This is a quite flexible method for face location.

(3) For more general applications, additional rules need to be put into the knowledge base to take into account faces which are tilted, turned and/or with glasses.

(4) Level 3 can be used for further face recognition by including more complex features.

(5) A syntactic method may be suitable for human face location based on the mosaic images.[15] Reconstruction of our face location algorithm to an algorithm using syntactic methods is not difficult.

Acknowledgement—This work was supported by a grant from Sumitomo Electric Industries.

REFERENCES

1. R. J. Baron, Mechanisms of human facial recognition, *Int. J. Man–Machine Studies* **15**, 137–178 (1981).
2. C. J. Wu and J. S. Huang, Human face profile recognition by computer, *Pattern Recognition* **23**, 255–259 (1990).
3. M. Turk and A. Pentland, Face recognition using eigenfaces, *Proc. IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 586–591, June (1991).
4. M. Kosugi, Human face identification using mosaic pattern and BPN, ACNN'91 (1991).
5. J. L. Perry, Human face recognition using a multi-layer perceptron, IJCNN'90-WASH, pp. II-413–II-416 (1990).
6. I. Craw, H. Ellis and J. R. Lishman, Automatic extraction of face features, *Pattern Recognition Lett.* **5**, 183–187 (1987).

7. M. Clark, Face recognition using a connectionist model, David Sarnoff Research Center and the University of Pennsylvania, pp. 1–12, May (1989).
8. L. D. Harmon *et al.*, Machine identification of human faces, *Pattern Recognition* **13**, 97–100 (1981).
9. M. Kirby and L. Sirovich, Application of the Karhunen–Loeve procedure for the characterization of human faces, *Trans. Pattern Analysis Mach. Intell.* **12**(1), 103–108 (1990).
10. V. Govindaraju, D. B. Sher, R. K. Srihari and S. N. Srihari, Locating human faces in newspaper photographs, *Proc. IEEE-CS Conf. of Computer Vision and Pattern Recognition*, San Diego, California, pp. 549–554 (1989).
11. V. Govindaraju, S. N. Srihari and D. B. Sher, A computational model for face location, *Proc. IEEE, 3rd Int. Conf. on Computer Vision*, Osaka, Japan, pp. 718–721 (1990).
12. V. Govindaraju, S. N. Srihari and D. Sher, A computational model for face location based on cognitive principles, *Proc. AAAI'92*, San Jose, California, pp. 350–355 (1992).
13. T. Agui, Y. Kokubo, H. Nagahashi and T. Nagao, Extraction of face regions from monochromatic photographs using neural networks, *Proc. Int. Conf. on Robotics*, Singapore, pp. CV18.8.1–CV18.8.5 (1992).
14. L. D. Harmon, The recognition of faces, *Scient. Am.* **229**, 71–82 (1973).
15. G. Z. Yang, On the knowledge-based pattern recognition using syntactic approach, *Pattern Recognition* **24**, 185–193 (1991).

About the Author—GUANGZHENG YANG received the B.E. and the M.S. degrees with Honours in electrical engineering from Leningrad Polytechnic Institute, Leningrad, U.S.S.R. He is now a Professor in the Department of Automation at the University of Science and Technology of China, Hefei, Anhui, China. From December 1991 to January 1993 he was a visiting scientist at the Beckman Institute, University of Illinois at Urbana–Champaign. His current research interests include pattern recognition, artificial intelligence, and computer vision.

About the Author—THOMAS S. HUANG received his B.S. degree in Electrical Engineering from National Taiwan University, Taipei, Taiwan, China; and his M.S. and Sc.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, Massachusetts. He was on the Faculty of the Department of Electrical Engineering at MIT from 1963 to 1973; and on the Faculty of the School of Electrical Engineering and Director of its Laboratory for Information and Signal Processing at Purdue University from 1973 to 1980. In 1980, he joined the University of Illinois at Urbana–Champaign, where he is now Professor of Electrical and Computer Engineering and Research Professor at the Coordinated Science Laboratory, and at the Beckman Institute. During his sabbatical leaves Dr Huang has worked at the MIT Lincoln Laboratory, the IBM Thomas J. Watson Research Center, and the Rheinishes Landes Museum in Bonn, West Germany, and held visiting Professor positions at the Swiss Institute of Technology in Zürich and Lausanne, University of Hannover in West Germany, and INRS-Telecommunications of the University of Quebec in Montreal, Canada. He has served as a consultant to numerous industrial firms and government agencies both in the U.S. and abroad. Dr Huang's professional interests lie in the broad area of information technology, especially the transmission and processing of multidimensional signals. He has published 11 books, and over 300 papers in network theory, digital filtering, image processing, and computer vision. He is a Fellow of IEEE, and the Optical Society of America; has received a Guggenheim Fellowship (1971–72), an A. V. Humboldt Foundation Senior U.S. Scientist Award (1976–77), and a Fellowship from the Japan Association for the Promotion of Science (1986). He received the IEEE Signal Processing Society's Technical Achievement Award in 1987 and the Society Award in 1991. He was a founding editor, and is currently an Area Editor, of the *International Journal of Computer Vision, Graphics, and Image Processing*; and Editor of the Springer Series in *Information Science* published by Springer Verlag.