System for the recognition of human faces

by M. S. Kamel
H. C. Shen
A. K. C. Wong
R. I. Campeanu

This paper describes a system for content-based retrieval of facial images from an image database. The system includes feature extraction based on expert-assisted feature selection, spatial feature measurement, feature and shape representation, feature information compression and organization, search procedures, and pattern-matching techniques. The system uses novel data structures to represent the extracted information. These structures include attributed graphs for representing local features and their relationships, n-tuple of mixed mode data, and highly compressed feature codes. For the retrieval phase, a knowledge-directed search technique that uses a hypothesis refinement approach extracts specific features for candidate identification and retrieval. The overall system, the components, and the methodology are described. The system has been implemented on an IBM Personal System/2® running Operating System/2®. Examples demonstrating the performance of the system are included.

Facial identification is one of the most challenging tasks for machine recognition. For humans it is an important social skill. Humans recognize faces even under different variations and alterations such as makeup, hair style, aging, and other physical effects. As for machine recognition of faces, it is still in an early stage. It is of interest because of its possible application in areas such as plastic surgery, security and personal checking systems, teleconferencing, and sign language.

There have been many attempts at developing systems to recognize faces. They can be classified into two categories: holistic approaches and analytic approaches.

The holistic methods emphasize the global properties of the form of a pattern. Some are based on neural networks,[1,2] others on the analysis of the isodensity lines from full facial images.[3] In the neural network approach of Kohonen et al.,[1] the network consisted of a multilayer of nodes with feedback, able to handle 80 × 40 input points. To evaluate the performance of their system, they used 10 classes of pictures, each of which consisted of five photographs of an individual taken from different angles, for training. The classifier was then able to recognize with reasonable accuracy a picture taken at angles not used in prior training. In WISCAD[2] the pattern recognition system was based on a digital learning net. To determine its effectiveness, 16 individuals were asked to face the camera in a face-on position with a variation of expressions allowed. As in the work of Kohonen et al., WISCAD requires many training instances consisting of different views of each individual who is to be recognized. In order

to achieve a 100 percent recognition rate for the group of 16 individuals, the users had to create up to 400 training pictures. The work of Sakaguchi et al.[3] uses the isodensity lines from full facial images. Their system was tested on 240 individuals. For each of them two images were taken, one for training and one for testing. The disadvantage of this system is that it requires images to be taken at a precise face-on position and under a special illuminating device. It can be concluded that because of special requirements or long training the holistic methods had rather limited success.

The analytic approaches, which concentrate on studying the spatial domain feature extraction, seem to have more practical value than the holistic methods. In these approaches specific facial features are extracted manually or automatically by an image processing system and stored in a database. A search method is then used to retrieve candidates from the database.

A number of studies have been conducted on the recognition of facial profiles.[4,5] The fiducial feature points extracted from the outline curve are represented in pattern vectors and used as in classical statistical pattern recognition. Although this method is quite effective in identifying up to a few hundred classes of faces, it is not clear whether or not these features are sufficient for a system that contains an even larger number of facial images. Another drawback is that face profile recognition does not satisfy the requirements of a practical automatic face identification system.

Although some attempts to automatically extract front view facial features have been made,[6,7] most of the existing systems are interactive and combine qualitative and quantitative features.[8–10] Such systems are in demand since police agencies require systems that are able to assist investigators in identifying individuals from a set of descriptions of qualitative features. One example is UHMFS,[8] which uses 21 attributes, each with five possible values, and nine geometric distances to characterize each individual. Another system is FRAME,[9] which holds a database of 1000 images of faces. Each face is associated with 50 parameters, including age, height, weight, some geometric measurements, and some descriptive parameters, rated by a group of trained judges. A third example is the system proposed by Riccia and Iserles,[10] which, apart from qualitative and quantitative data about the individual, also includes information on the type of crime committed.

A common drawback for all of these systems is that, although they include geometric measurements, they are unable to take into consideration perspective variations and the invariant properties of the interfeature point distances. Since the searched individual is most often photographed in positions that are different from the face-on position, the consideration of the perspective variations is of great importance for any practical facial recognition system.

Our system belongs to the group of analytic approaches using front view facial images. Unlike the systems described above, we have a special feature extraction method that implies little, if any, user expertise. We have also created new transformation and matching techniques that will consider perspective and invariant properties and thus allow for the recognition of the faces with posture variation. In the design and the implementation of our system, the following tasks have been identified:

- Determination of pertinent facial features that have been effectively used by investigators
- Identification of image processing and analysis problems and methodologies related to the extraction of those pertinent features
- Conceptualization and building of a knowledge-based prototype system in which the extracted features could be encoded and used for screening, retrieval, and identification of candidates (taking into consideration the functional and operational characteristics of a fully automated and intelligent system)

Image feature selection

To present a fuller picture of how and what features should be selected for candidate screening and identification, we first address two notions, namely, feature categories and feature reliability.

Feature categories. There are two feature categories, qualitative and quantitative.

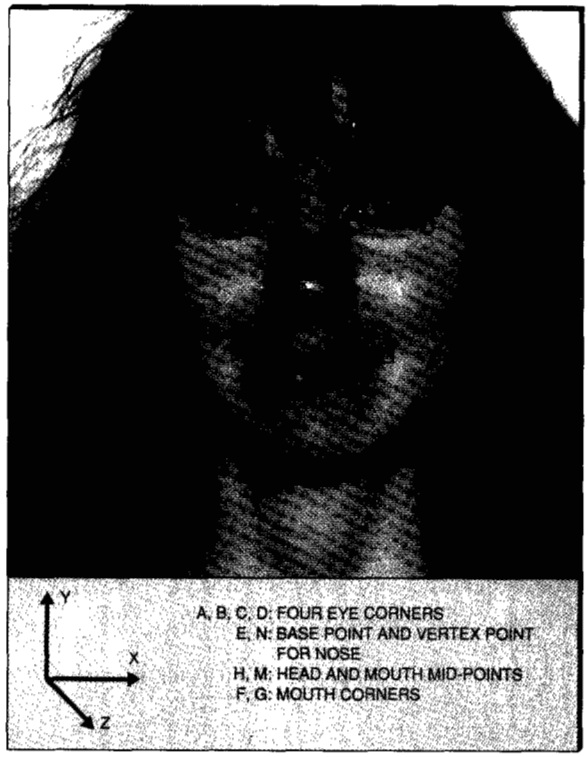
Qualitative features. Qualitative features are often used by humans. These features provide significant information for retrieval and screening. When applied to machine vision, they are not easily and consistently acquired, and their effective-

ness highly depends on the robustness of the image processing and pattern recognition systems. At this stage, we use only features that can be acquired and recognized in the least ambiguous fashion. As feature extraction progresses, more "descriptive" features represented in quantitative or symbolic forms, or both, will be added. Examples of features used at this stage are the position of the ears relative to eye and nose levels, the size of the nostrils, the shape of the chin, and the mean curvature of the eyebrows and their distance to the eyes. Nevertheless, these features are not always visible in an image because the ears might be hidden by hair and the chin by a beard, and the shape of the eyebrows is sometimes affected by facial expressions. But once precisely extracted, they can be effectively used for both candidate screening and identification. The co-occurrence or absence of these features would significantly help to effectuate the screening process and enhance reliability.

Quantitative features. For identification, spatial relation of distinct facial features usually furnishes highly redundant information. In our early study,[11] we found that the configurations of interfeature distances are extremely effective for subject comparison and identification. The number of interfeature distances that proved to be important in face identification decreased during the evolution of our project. Initially, based on the existing literature, our work considered 14 interfeature distances, which implied 23 feature points. At the end of this project seven feature points were proven sufficient for the identification of a face out of a group of over 80 faces. However, it is possible that for larger databases the number of important feature points could be greater.

Since the same posture of a subject cannot be easily obtained in a photograph or a live image in real practice, perspective invariant features have to be developed to ensure the robustness of an automated system. To meet this requirement, a new feature configuration "invariant" to perspective transformation, developed for recognizing perspective invariant features in three-dimensional vision, was modified and used for this purpose. On this configuration, a "cross ratio" that is independent of both spatial and facial expressions can be defined. Based on the deviation of certain measurements from symmetry, the rotation of the head relative to a vertical plane can be estimated, and the projection of a three-dimensional feature configuration on an imaginary plane can be recovered and used for identification. The cross ratio together with random graphs[12] can also be used to reduce the search space and enhance effectiveness during the retrieval process. Hence, in an image, essential human face features that usually fluctuate with variations in posture can be recovered. Experiments have demonstrated the effective use of such feature configurations. Figure 1 shows the set of facial features that collectively defines a configuration possessing the above-described characteristics. Their extraction and measurement from an image are relatively easy to accomplish. They can be used for matching the subject in an image with potential candidates retrieved from the database, if the orientation of the head of the subject or the candidate facing the camera is less than 45 degrees.

Figure 1 A configuration of a set of features



One of the important steps in the processing of our configuration of interfeature distances is an adequate normalization. This normalization ensures that the distance of the subject to the cam-

era, or the size of the digitized photo, plays no role in the recognition process.

Feature reliability. The next notion to be addressed is the reliability of the features used. The major concern for qualitative features is their acquisition, but for quantitative features it is their reliability that can be categorized according to their effectiveness:

First grade: Spatial (rotation, translation, or scaling) and facial expression invariant

Second grade: Spatial invariant but affected by facial expression, or spatial dependent but not affected by facial expression

As shown in Figure 1, almost all of these features are described by horizontal (X direction) and vertical (Y direction) linear measurements. Since in most cases, the human head posture is captured under rotations about the Y-axis (Figure 1), the vertical measurements will be quite invariant from one photo to another. The horizontal ones are the most distorted by the Y-axis rotation. Yet a very effective method to recover this information is presented in the next section. Thus, the distances between the four eye corners A, B, C, and D and those between the head midpoint H and the mouth midpoint M (or the nose base point E) can be viewed as features belonging to the first grade. The width of the mouth and the height of the lips or that of the eyes (not considered at this stage of our work) fall into the second grade.

Feature measurements

This section presents the extraction of the feature points, their use in the derivation of a set of invariant interfeature distances, and the determination of some qualitative features.

The following three assumptions were at the basis of our measurement process:

1. The rotation of the head about the X-axis is not significant.
2. The four corners of the eyes (A, B, C, D) are colinear to form a straight line (D_1).
3. The line (D_2) passing through the mouth corners F and G and (D_1) will form a plane $\langle P \rangle$.

The first assumption is in fact related to the second: If the rotation of the head about the X-axis is significant, the line (D_1) will be curved. Although the derivation of a set of feature distances invariant for rotations about both the X-axis and Y-axis is a useful extension of our present work, we found that all of the pictures that we studied could be considered to be in agreement with the above assumptions.

Feature points measurements. The user interface for data capture contains pull-down menus, text windows, a window containing a diagram of a generic face sketch with the feature points, the highlighting of the area containing the feature to be captured, etc.

In the present work only seven feature points (A, B, C, D, E, F, and G in Figure 1) are extracted from each picture. The extraction is semiautomated. The process of extracting feature points involves three stages:

1. Either the image of a human subject or a photo of the subject is digitized by a CCD (charge coupled device) camera. Each digitized image is stored in an image database, and it is displayed for capture of the features. Next to the image the system presents a features list window containing the feature names and X-Y coordinates.
2. The user indicates which feature is to be located on the image by clicking a mouse on the appropriate feature name (in the feature list window). As a result, a cross-hair marker is positioned in the vicinity of the chosen feature point.
3. The operator clicks the mouse, after having established the exact position of the feature point. At that moment the X- and Y-coordinates are entered into the feature list window.

The process is repeated for all seven points of the features. At any time the operator can return to an already-located feature in the feature list window and can relocate that feature if necessary.

Invariant interfeature distances. After capture of the point features, the line (D_1) is built by using the least squares technique, and the four eye corners are projected back on this line. Second, from assumptions 2 and 3, four new lines (R_i), (R_o), (L_i), and (L_o) can be drawn (Figure 2), and their intersections at points R and S then belong to the

Figure 2 Determination of feature points for recognition

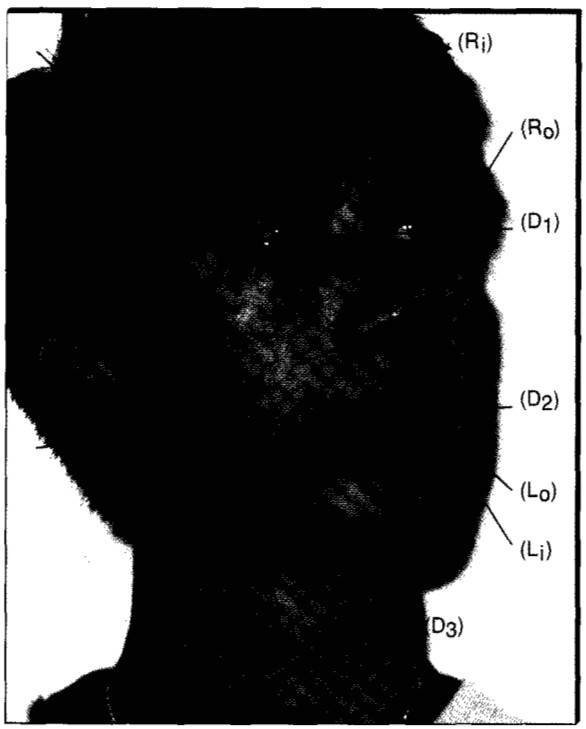
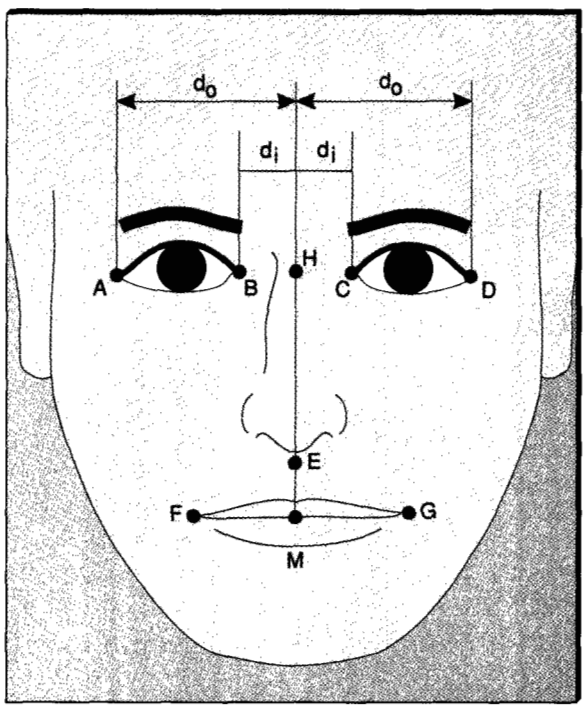


Figure 3 Feature points and interface distance used in the recognition process



plane $\langle P \rangle$. Consequently, the axis of symmetry (D_3) for the human face that can be deduced from R and S will intersect (D_1) and (D_2) to give the head midpoint H as well as the mouth midpoint M.

Given that the "cross ratio" of any four points on a line is completely independent of any perspective transformation, the one $R(A, B, C, D) = (d(A, C)*d(B, D)) / (d(A, D)*d(B, C))$ on line (D_1) will constitute an invariant feature. In this notation $d(A, C)$ is the distance between points A and C, $d(B, D)$ the distance between points B and D, etc.

Even in reality, the left and right parts of the human face are not perfectly symmetric. However, the following hypothesis is adopted in our measurement estimation (Figure 3):

$$\begin{cases} d(H, B) = d(H, C) = d_i \text{ and} \\ d(H, A) = d(H, D) = d_o \end{cases}$$

This implies that

$$R(A, B, C, D) = (d_i + d_o)^2 / 4 \times d_i \times d_o$$
$$= (1 + \beta)^2 / 4 \times \beta$$

where $\beta = d_i/d_o$, will provide a first estimate β_1 of β. Moreover, $R(A, B, H, C) = R(D, C, H, B) = 2 \times d_o/(d_o + d_i) = 2/(1 + \beta)$ implies a second estimate β_2 that can be expressed as

$$\beta_2 = (1/R(A, B, H, C) + 1/R(D, C, H, B) - 1)$$
$$\Rightarrow \beta = (\beta_1 + \beta_2)/2$$

When the human head is turned to one side (for example, to the right and less than 45 degrees), the distance $d(A, C)$ will be much more precise than $d(B, D)$ and will nearly stay invariant as shown in Table 1. Thus, only corrections on points B and D are necessary to recover the original distances $d(C, D')$ and $d(B', D')$ (all of the others can be easily deduced from these two distances by applying symmetry on the four eye corners).

Table 1 Invariant properties of various interfeature point distance measures (all distances are given in millimeters)

AC	AD	BD	FG	HM	HE	Position	Candidate
56.06	83.02	56.01	44.00	50.50	31.46	FrtB	Andrew
56.07	83.08	56.07	44.03	50.48	31.45	FrtA	Andrew
54.08	73.17	44.17	39.20	52.28	34.57	T45B	Andrew
56.08	83.08	56.08	44.51	48.96	32.37	T45A	Andrew
46.00	66.00	45.00	37.00	37.50	25.52	FrtB	BetMca
46.19	67.00	46.19	37.56	37.22	25.33	FrtA	BetMca
45.00	64.01	40.00	34.00	40.51	26.48	T45B	BetMca
45.90	69.01	45.90	36.66	38.99	25.48	T45A	BetMca
52.03	74.01	51.99	38.00	44.00	28.00	FrtB	KimNgu
52.03	74.04	52.03	38.02	43.99	27.99	FrtA	KimNgu
49.09	66.07	43.05	32.00	45.63	29.30	T45B	KimNgu
50.42	72.11	50.42	34.93	43.64	28.02	T45A	KimNgu
52.88	76.10	52.08	36.06	50.33	34.30	FrtB	LindPat
53.04	76.90	53.04	36.44	50.07	34.12	FrtA	LindPat
51.07	70.03	43.98	35.00	52.26	34.48	T30B	LindPat
52.50	77.12	52.50	38.54	49.75	32.82	T30A	LindPat
78.27	110.16	78.04	65.07	80.82	58.10	FrtB	Chan
78.32	110.39	78.32	65.20	80.74	58.04	FrtA	Chan
77.98	108.02	75.04	59.03	81.37	54.90	T30B	Chan
78.62	110.96	78.62	60.64	80.28	54.16	T30A	Chan
72.00	99.00	71.01	51.01	62.76	43.29	FrtB	Lian
72.22	99.99	72.22	51.52	62.45	43.08	FrtA	Lian
67.96	89.14	59.18	44.18	63.59	44.10	T45B	Lian
70.08	97.92	70.08	48.53	60.60	42.03	T45A	Lian

FrtB, FrtA: Frontal picture before and after applying the cross ratio correction.
T45B, T45A (T30B, T30A): Turning at 45 (30) degrees before and after applying the cross ratio correction.

Let D′, B′ be the original points so that

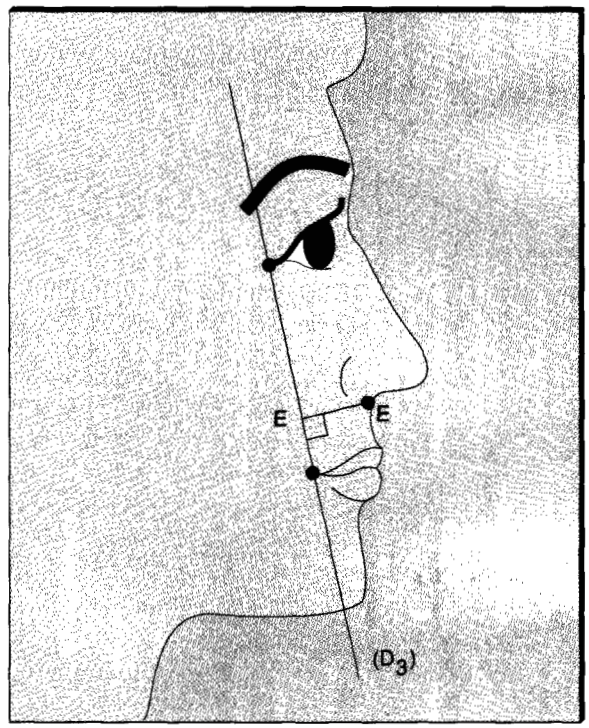
$$\begin{cases} d(H, D') = d(H, D) + X_d, \\ d(B', D') = d(B, D) + X_d - X_b \end{cases}$$

then from

$$d(H, D')/d(B', D') = d_o/(d_o + d_i) = 1/(1 + \beta) = \alpha \approx (d(H, D) + X_d)/(d(B, D) + X_d - X_b)$$

and from $d(B', D') = d(B, D) + X_d - X_b = d(A, C)$ it follows that a system of two equations of two unknowns X_d and X_b

Figure 4 The actual eye-nose distance



$$\begin{cases} (1 - \alpha)X_d + \alpha X_b \approx \alpha d(B, D) - d(H, D) \\ X_d - X_b \approx d(A, C) - d(B, D) \end{cases}$$

can be solved to infer the original distances $d(C, D')$ and $d(B', D')$ (for $d(C, D') = d(A, B') = d(A, B) + X_b$).

Now, let $k = d(A, D')/d(A, D)$; then the original distance $d(F', G')$ can be expressed as

$$d(F', G') = k \times d(F, G)$$

Because the base point E under the nose is not necessary on the plane $\langle P \rangle$ determined by A, B, C, D, F, and G, the real distance $d(H, E')$ is obtained by projecting $d(H, E)$ on the symmetry axis (D_3) (Figure 4).

In summary, the four most reliable quantitative features consist of $d(A, C')$, $d(A, D')$, $d(H, E')$, and $d(H, M)$. $d(H, M)$ would change little; even expressions of smiling or the mouth opening

Table 2 The qualitative features Cut_1 and Cut_2 and their assigned values

	Cut_1	Cut_2
Information not available	−1	−1
Not being cut	0	0
Being cut by $\langle P_2 \rangle$	0	1
Being cut by $\langle P_1 \rangle$ and $\langle P_2 \rangle$	1	1
Being cut by $\langle P_1 \rangle$	1	0
Being cut and the lower part is greater than the upper one	0	2

would not affect the components of F and G along the Y-axis direction. The mouth width is more sensitive to facial expression but is still taken into consideration during the matching process since it forms a very typical feature for each candidate.

Qualitative features. In the present work only two qualitative features were considered: the size of the nostrils and the position of the ears. Although the size of the nostrils can be simply defined as large, medium, or small, the position of the ears is determined in a more elaborate way.

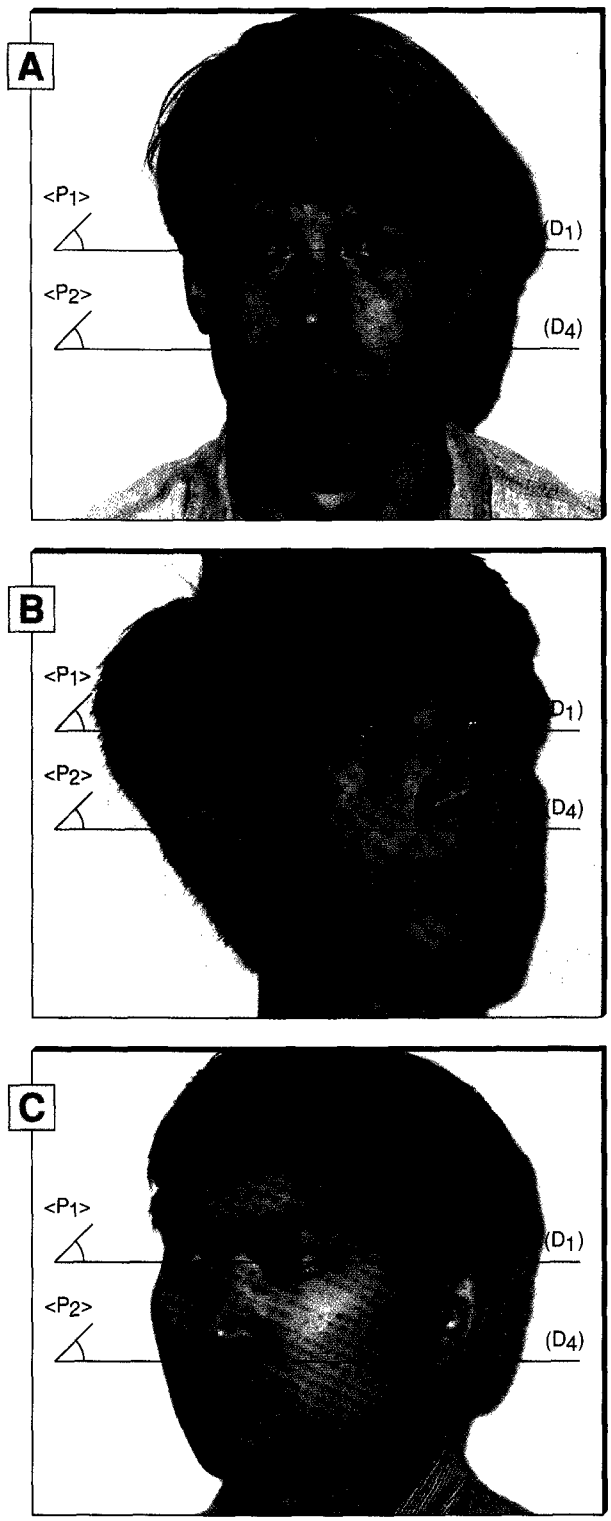
The ears were visible in most of our photos, and this qualitative feature proved to be extremely useful. The ear position can be defined as related to the line (D_1) through eye corners and to the line (D_4) through the nose base-point E having the same direction as (D_1). In general, the ears will be one of the following:

1. Located between the two planes $\langle P_1 \rangle$ containing (D_1) and $\langle P_2 \rangle$ containing (D_4) (both $\langle P_1 \rangle$ and $\langle P_2 \rangle$ are perpendicular to the symmetry axis (D_3))
2. Cut by these two planes
3. Cut by one of these two planes

Consequently, creation of two new qualitative attributes is required: "Cut_1" (cut by $\langle P_1 \rangle$) and "Cut_2" (cut by $\langle P_2 \rangle$), with the assigned values given in Table 2.

The three photos in Figure 5 therefore have the following values assigned to Cut_1 and Cut_2: A has (1 0), B has (0 2), and C has (1 1). In the present system the user has to judge and then manually enter into the database all of the qualitative feature values.

Figure 5 The qualitative features Cut_1 and Cut_2



Similarity measurements

As the exposure distance of a subject relative to the camera might vary considerably, adequate feature "normalization" is vital to ensure an effective similarity measurement. This normalization, together with the perspective of the chosen features, will also ensure that the photo sizes, or the positioning angle of the photo, will not influence the recognition process.

Feature normalization. This step is of paramount importance in compensating the scaling variations between quantitative features from two postures of the same candidate.

Let F_i, i = 1, . . . , 5 be the set of quantitative features extracted and processed according to the techniques described in the previous section on feature measurements. The common ways to achieve the normalization rely on either:

1. Selecting a reliable feature for the normalization basis, i.e., if $F_k \neq 0$ is assumed to be reliable (see feature reliability), then $\{F_i/F_k\}$ will constitute a set of normalized features that "theoretically" are scale-invariant (the normalization basis F_k ought to be quite precise, otherwise it will affect the accuracy of each F_i/F_k; $k \times i$)
2. Dividing $\{F_i\}$ into two subsets and processing the normalization separately as in 1, i.e., $\{F_i\} = \{F_{hi}\} \cup \{F_{vi}\}$ and the normalized set will be $\{F_{hi}/F_{hr}\} \cup \{F_{vi}/F_{vs}\}$, where $\{F_{hi}\}$, $\{F_{vi}\}$ are the horizontal and vertical quantitative features, respectively, and $\{F_{hr}\}$ and $\{F_{vs}\}$ are considered as some reliable features among the horizontal and vertical ones, respectively.

However, as shown in the next subsection, the similarity measurements obtained from the above methods will not be promising. Since after such normalization some precious information will be lost, the so-called "reliable" features F_k (case i) or F_{hr} and F_{vs} (case ii) now all becoming unitary can no longer be used for feature comparison. In other words, the number of meaningful quantitative features will be reduced to four (case i) or to three (case ii). Therefore, in order to preserve the five original quantitative features as well as to get around the critical selection of a reliable feature, the vector concept is adopted, i.e., quantitative features extracted from a human face will be represented by a vector:

$$F = (F_1, F_2, \ldots, F_5)$$

and the corresponding normalized vector, which is scale-invariant, will be $N = F / \|F\|$.

Similarity measure. Let $N_1, N_2 \in \{N_i\}$ (set of quantitative feature vectors for all subjects stored in the database) represent candidates 1 and 2. Candidate 2 is said to be the most similar to candidate 1 if and only if

$$\|N_1 - N_2\| = min\{\|N_1 - N_i\|, i \neq 1\}$$

Since all norms are equivalents in R^5 and especially for saving in CPU time, we replace the Euclidian norm by the city block one, i.e.,

$$\|N_1 - N_2\| = \Sigma \, |n_{1i} - n_{2i}|$$

One should observe that this similarity measure is used only for "ranking" the degree of resemblance between different subjects (already classified into a resemblance class by applying the subgroup concepts described in the next section) and not for the purpose of screening.

Candidate retrieval and identification

The two major steps for the retrieval and identification of human faces are, respectively, the "subgroup concept" screening and the resemblance ranking.

Subgroup concept screening. Let S_i be a subgroup that is defined as a set formed by p feature vectors N_k; $k = 1, \ldots, p$ satisfying the following criteria:

for any $N_l = (n_{l1}, \ldots, n_{l5})$ which does not belong to $S_i = N_1, N_2, \ldots, N_p$ then

$$|n_{li} - n_i| > max\,\{|n_{ki} - n_i|\};\ k = 1, \ldots, p$$

This implies that the p feature vectors N_k ($k = 1, \ldots, p$) forming S_i will have the i^{th} component classified as "the top p" closest to n_i.

Consequently, by building each S_i, the screening process that reduces the number of human faces to be matched will take into account each quantitative feature separately. This subgroup concept can be extended to the discrete measurements, in particular to the "qualitative features."

For instance, if n_6 and n_7 are the two qualitative features *Cut*_1 and *Cut*_2, the feature vector will become $N = (n_1, \ldots, n_5, n_6, n_7)$ while the corresponding subgroup S_6 (or S_7) will be defined as above if n_6 (or n_7) is different from -1 and equal to the whole set of feature vectors (complete set of subjects stored in the database) if n_6 (or n_7) is -1. (We recall that a negative value assigned to the qualitative *Cut*_1 or *Cut*_2, or both, is equivalent to unavailable information.)

Now, let $S = \cap\, Si(i = 1, \ldots, 7)$. This intersection set is called the "resemblance class" for the candidate N (or the feature vector N) and has, in general, an order less than p. Sometimes S might be empty since two subgroups S_i, S_j generated from two different features will not necessarily contain the same element N_k, $(k = 1, \ldots, p)$. Should this happen, the reason most probably is that either the number of p subjects to identify with the candidate N is too small or the similarity between the test candidate and the other subjects is negligible, and at the same time, these other subjects are very different from each other.

Resemblance ranking. This final phase aims at evaluating the degree of similarity between the test candidate and those classified in the resemblance class. The similarity measure will be the city block norm applied to the vector quantitative features (presented in the subsection on similarity measure) by means of excluding the last two qualitative components. The ranking process will make use of some simple sorting algorithms, such as bubble sorting. Finally, it is understood that the number of rankings has to be less than the order of the resemblance class which is less than p.

Feature information organization and image database

Feature information needs to be stored in a database for the purpose of retrieval. The way in which this information is organized in a database depends on the levels of retrieval and the type of information required.

The levels of retrieval that may be allowed are the following:

- Retrieval by contextual profile—In this retrieval, information about the candidate is provided to the system in text form (descriptive or approximate measurements).
- Retrieval by similarity of a given sample—Here the information presented to the system is an image.
- Retrieval by a sketch—In this level of retrieval a sketch of the candidate is provided. The information is based on a graphical rather than a pictorial image.

The following types of information may be required from the retrieval:

- The detailed images that satisfy the query
- Only contextual information if an exact match is found
- Only statistical information on possible match, which may be used to refine the query
- Combinations of the above

Each of the above will have some impact on the database used for the system. Different approaches for designing image databases to accommodate some of the above requirements have been proposed.[13,14]

In this paper we shall present only data related to the retrieval by similarity of a given sample. The various parts of our project (data capture, database system, and face recognition) were integrated into a single shell program, which has two available modes, either of which can be enabled by an operator through the use of pull-down menus:

- The build database mode
- The face recognition mode

As the retrieval is based on the similarity of a given sample, the face recognition mode contains the same data capture process as the build database mode, after which the shell passes the features information and the control to the face recognition subsystem, which in turn accesses the database subsystem.

SQL interface to database. Structured Query Language (SQL) provides an effective interface to databases. Access to the image data can be performed by low-level interface functions. The complete image database consists of the files of images and a number of significant tables. A brief summary is given to outline the configuration of the database.

Figure 6 Database interface routines

```
struct  feature_data_struct  {
   id_type              id;
   string               name;
    feature_type1       f1;
    feature_type2       f2;
     ...
}

get_a_file_name    (id,  image_data_struct)
put_a_file_name  (id,  name)
id_type     id;
string      name;
{
     ...
    UPDATE image_file_name
   SET file_name = name
  WHERE ID = id
     ...
}

get_a_PF  (id,  pf_data_struct)
put_a_PF  (id,  pf_data_struct)

get_a_FM  (id,  fm_data_struct)
put_a_FM  (id,  fm_data_struct)

get_a_QF  (id,  qf_data_struct)
put_a_QF  (id,  qf_data_struct)
```

1. Image files: stored as grey level
2. Image file name table: associates a unique identifier and the image file particulars
3. Point feature table: associates an identifier and the point features extracted from the corresponding image
4. Feature (quantitative) measurements table: associates an identifier and the calculated feature measurements
5. Feature (qualitative) measurements table: associates an identifier and the qualitative feature measures

Examples of these tables are shown in Figures 6 and 7.

Implementation and performance

The proposed system has been developed and implemented on an IBM Personal System/2* Model 80 computer running under Operating System/2* (OS/2*) EE 1.1. Images are captured by a JOVIAN** digitization card and the companion VU** software package. Unfortunately, only the DOS (Personal Computer Disk Operating System) version is available to date. This means swapping between OS/2 and DOS is necessary at present. Images are digitized into 320 × 240 pixels of 64 grey levels. Using the VU software package, the captured images can be modified, edited, and displayed. In our system, only the capture and display facilities are used.

As already explained, the program has two available modes. When the build database mode is chosen, the image is captured and displayed, and qualitative and quantitative features are extracted and stored in the database. When the face recognition mode is chosen, the search subject image is captured, the features extracted, and the screening and resemblance ranking procedures are executed. In this case, images will be retrieved from the database based on the level of retrieval that is required and the type of feature measures that are supplied.

The program typically produces the 10 most likely matches (the number 10 can be modified by the user), based on the city block norm defined similarity measure. The system also allows the user to request additional candidates. Each search presents a list of image file names of the potential candidates in descending order of the similarity measures (also shown in the list). The user can then display the search subject image and the candidate images and make a final decision.

Our tests were executed with a set of 84 images: 45 were obtained by digitizing existing photos, while 40 were obtained by capturing and digitizing the images of a number of subjects. For each of these subjects a number of pictures were taken, each corresponding to a different posture. The tests were all based on the "hold-one-out" idea, i.e., one of the 84 images was the search subject, and the database searched had 83 images. In the list of the 10 candidates that match the subject, the similarity measures varied, depending on the search subject. In some cases all of the similarity measures were relatively small; in others the similarity measures for some of the candidates were fairly large (i.e., they were very unlikely candidates).

We noticed that if for a certain search subject the database contained images of the same person but in different postures, the search always put the corresponding images in the candidates list. In 95 percent of cases the person was in the top four on the similarity list, in 86 percent of the cases the

person was in the top two on the list, and in 66 percent of the cases, the person was number one on the list.

Our review of existing systems shows that no other project using an analytic method contained perspective invariant features, and therefore, there are no similar results against which to compare our data. The only other face recognition systems allowing for perspective variation were of the holistic type.[1-3] As shown in the introduction, in order to achieve a good recognition rate, the holistic systems require extremely long training or special image capture conditions. Our analytic method is by comparison much simpler.

The performance of our system was achieved by the use of the cross ratio correction as explained in the section on the feature measurements. Table 1 showed that the inclusion of this correction makes the relevant interfeature point distances nearly rotation-invariant.

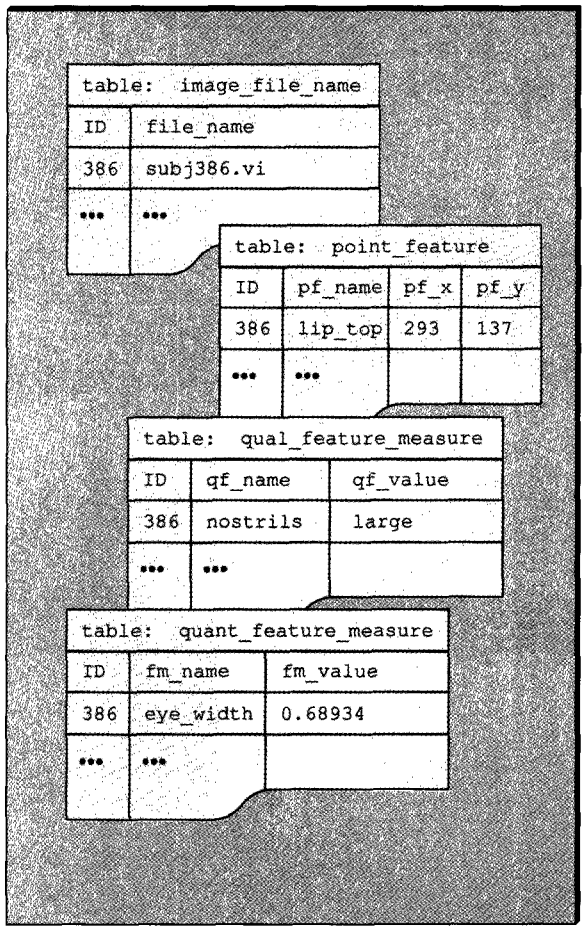
Concluding remarks

The work reported in this paper can be considered as a first phase of a comprehensive study on feature extraction and recognition for content-based retrieval in image databases. Our work demonstrates that the technical problems associated with this system can be very well-defined. Very innovative ideas on recognizing subjects with different posture and perspective views have been proposed and proved feasible in real-world experimentation. The overall system configurations have been conceived and put together into an integrated system.

One of the areas in which we have made considerable progress is the user-machine interface. The present stage, although only semiautomated, provides the user with help in the form of menus, text, and image windows and highlighting of the area where the feature points are to be extracted. The user is guided through the steps associated with capture of features. Much more work is required in order to provide the system with a fully automated feature-capture mechanism.

Although we introduced a number of perspective invariant quantitative features, we are still limited to facial images that have a rotation about the Y-axis. If the face of the search subject is significantly rotated about the X-axis, we obtain search

Figure 7 Database table configuration



results that are worse than the percentages reported in the previous section. More work is required to provide a new set of quantitative features that is perspective-invariant relative to rotation about both axes.

Research in the present phase is still dealing with a relatively small set of data though it is sufficient to demonstrate the technical capability of the system. However, having this methodology applicable to a very large database is our ultimate goal. Hence, the notions of further data compression, synthesis, and organization have to be addressed.

For candidate retrieval, the local facial features, their parametric representation, the perspective-invariant cross ratios of feature configurations, and the canonical feature configuration repre-

sented in the form of attributed graphs are all important. Ideally the compressed information should be represented in a structural pattern. From our previous research experience and our preliminary study, the structural patterns can be represented in two essential forms:

1. Tabulated features of mixed types (or n-tuples of mixed mode data),[15–17] i.e., a mixed set of ordinal (continuous or discrete) and nominal data. For instance, the cross ratio of the perspective-invariant feature configuration is a real number; the Fourier descriptor of the chin contour is a subset of real numbers, and its categorization is nominal data; the type of noses or the qualitative descriptor of the shape of an eye is nominal data and so forth.
2. Relational structure of facial features such as the perspective-invariant feature configuration. This configuration can be represented in the form of attributed graphs.[18,19] On this formal structure, special transformation can be used to transform its spatial distance values for rotational adjustment. Such an approach is also known as a knowledge-directed search based on hypothesis refinement that has been effectively applied to a robust three-dimensional vision system.[19]

When enormous amounts of data are involved, further data synthesis and compression are necessary. At the present stage, our existing database is still relatively small. However, special data synthesis and compression have to be considered to give a technological backup to the existing data structures as well as recognition, inference, and retrieval systems. In view of this, the following methodologies that have been successfully applied to other real-world problems are proposed:

1. The inductive learning on multivariate mixed-mode data based on event covering for synthesizing n-tuple into categories will be considered. This will help to reduce the search in the screening process. The event-covering algorithm is able to detect statistically interdependent patterns inherent in a set of ordered or unordered observed events.[20] And, for this reason, it is able to detect patterns in derived events that may not be directly observable. The ratios of certain measurements is an example. Hence, this method can be used in unsupervised classification for the autonomous organization of pattern categories.

   The algorithm can be divided into three phases:

   (a) Detection of underlying patterns in the observed or derived events
   (b) Construction of probabilistic rules or patterns so as to reveal more explicitly the underlying features and properties of the observed data
   (c) Decision-making based on the generated rules

   The algorithm has been tested with simulated as well as real-life data in both supervised and unsupervised learning tasks involving ordered or unordered and mixed-mode data. It renders excellent results.

2. To synthesize relational information represented in the form of attributed graphs, the structural pattern approach based on a random graph[12,21] will be adopted. By random graph, we mean a probability structure of representing an ensemble of graphs that share structural and contextual similarity. A random graph can be considered as the synthesis of such an ensemble, which in turn could be considered as the outcomes of the random graph. A random graph consists of a set of random vertices together with a family of random arcs whose occurrences are conditioned by the attributed values of their incident vertices. Associated with each random vertex is a probability distribution of its attributed values, and corresponding to each random arc is a conditional probability distribution of the arc attributed values given the values of its incident vertices. A random graph can thus be used to represent the synthesis of the canonical forms of feature configurations (including the possible inclusion of feature attribute values).

   With this data structure and a random graph synthesis algorithm, attributed graphs that share similar structural and contextual patterns will be synthesized into random graphs. Thus, when an attributed graph is to be categorized, it will be matched with a much smaller set of random graphs rather than with all attributed graph representations of all candidates in the entire database.

When applied to retrieval problems, this approach will be integrated with the cross-ratio screening. For those cases that fall within a distance threshold of the cross ratios as well as those being screened in by the event-covering methods, the search can be further narrowed by comparing their attributed graph representation with only the subset of random graphs. Once a subset of optimal graph mappings of the attributed graph to a subset of random graphs is found, the next level of random graphs or attributed graphs can be retrieved for further screening until the best subset of attributed graphs in the database is identified.

Future research, as we conceive it, is to consolidate each of the identified methodologies. For large image databases an important element will be the automated extraction of facial features. More data have to be collected and entered into the database system so that the second phase of data compression and synthesis research can commence.

Acknowledgment

This work was supported in part by a research contract from the IBM Toronto Laboratory. The authors would also like to acknowledge the development effort of the following graduate students at the University of Waterloo: Keith Chan, Tong-Minh Hong, Lian Guan, Bruce McArthur, Glen Newton, Kim Nguyen, and Queintin Tang.

*Trademark or registered trademark of International Business Machines Corporation.

**Trademark or registered trademark of Jovian Logic Corporation.

Cited references

1. T. Kohonen, P. Lehtio, J. Rovamo, J. Hyvarine, K. Bry, and L. Vainio, "A Principle of Neural Associative Memory," *Neuroscience* **2**, 1065–1076 (1977).
2. T. J. Stonham, "Practical Face Recognition and Verification with WISCAD," *Aspects of Face Processing*, H. D. Ellis, M. A. Jeeves, F. Newcomve, and A. Young, Editors, Nijhoff, Dordrecht, Boston (1986), pp. 426–441.
3. T. Sakaguchi, O. Nakamura, and T. Minami, "Personal Identification Through Image Using Isodensity Lines," *SPIE* **1199**, Part 2, 643–654 (1989).
4. G. J. Kaufman, Jr., and K. J. Breeding, "The Automatic Recognition of Human Faces from Profile Silhouettes," *IEEE Transactions on Systems, Man and Cybernetics* **SMC-6**, 113–121 (February 1976).
5. L. D. Haron, M. K. Khan, and P. F. Ramig, "Machine Identification of Human Faces," *Pattern Recognition* **13**, No. 2, 97–110 (1981).
6. K. K. Tsui, *Computer Recognition of Human Faces*, Ph.D. thesis, School of Electrical Engineering, University of Sydney, Australia (June 1985).
7. M. Nixon, *Proceedings of SPIE, International Society of Optical Engineering* **575** (1985), pp. 279–285.
8. G. W. Batten, Jr. and B. T. Rhodes, Jr., "UHMFS: The University of Houston MUG File System," *Proceedings of the 1978 Carnahan Conference on Crime Counter Measures*, Kentucky (May 1978), pp. 15–26.
9. J. W. Shepherd, "An Interactive Computer System for Retrieving Faces," *Aspects of Face Processing*, H. D. Ellis, M. A. Jeeves, F. Newcomve, and A. Young, Editors, Nijhoff, Dordrecht, Boston (1986), pp. 398–409.
10. G. D. Riccia and A. Iserles, "Automatic Identification of Pictures of Human Faces," *Proceedings of the 1977 Carnahan Conference on Crime Counter Measures*, Kentucky (May 1978), pp. 145–148.
11. A. K. C. Wong, M. S. Kamel, and H. C. Shen, "Feature Extraction and Recognition Techniques for Content Based Retrieval in Image Databases," *Part I: A Semiautomated System*, Progress Reports I and II, IBM Canada Laboratory, Toronto (1989).
12. A. K. C. Wong, J. Constant, and M. L. You, "Random Graphs," *Syntactic and Structural Pattern Recognition—Fundamentals, Advances, and Applications*, Vol. 7, H. Bunke and A. Sanfeliu, Editors, World Scientific Publishing Company Pte. Ltd., Singapore (1990), pp. 197–234.
13. T. Takao, S. Itoh, and J. Iisaka, "An Image-Oriented Database System," *Database Techniques for Pictorial Applications*, A. Blaser, Editor, Springer-Verlag, Berlin (1989), pp. 527–538.
14. H. Tamura and N. Yokoya, "Image Database Systems: A Survey," *Pattern Recognition* **17**, No. 1, 29–43 (1984).
15. K. C. C. Chan and A. K. C. Wong, *Automated Acquisition of Probabilistic Knowledge from Noisy Training Instances*, Report 156-M-880426, Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada (1988).
16. A. K. C. Wong and D. K. Y. Chiu, "Synthesizing Statistical Knowledge from Incomplete Mixed-Mode Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-9**, No. 6, 796–805 (1987).
17. A. K. C. Wong and K. C. Chan, "Learning from Examples in the Presence of Uncertainty," *Proceedings of the International Computer Science Conference '88: AI Theory and Applications*, Hong Kong (1988), pp. 369–376.
18. A. K. C. Wong and S. W. Lu, "Recognition and Knowledge Synthesis of 3-D Objects Based on Attributed Hypergraphs," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-11**, No. 3, 279–290 (1989).
19. K. D. Rueb and A. K. C. Wong, "Visual Part Identification and Location in a Robot Workcell," *International Journal of Machine Tools and Manufacture, Special Supplement on Robotics and Artificial Intelligence* **28**, No. 3, 235–249 (1988).
20. F. A. Akinniyi, A. K. C. Wong, and D. Stacey, "A New Algorithm for Graph Monomorphism Based on the Projections of the Product Graph," *IEEE Transactions on Systems, Man, and Cybernetics* **16**, No. 5, 740–751 (1986).
21. A. K. C. Wong and M. L. You, "Entropy and Distance of Random Graphs with Application to Structural Pattern Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-7**, No. 5, 599–609 (September 1985).

Accepted for publication October 20, 1992.

Mohamed S. Kamel *PAMI Group, Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.* Dr. Kamel received his B.Sc. (Hons) degree in electrical engineering from the University of Alexandria, Egypt, in 1970, his M.Sc. degree in computation from McMaster University, Hamilton, Ontario, in 1974, and his Ph.D. degree in computer science from the University of Toronto in 1981. From 1980 to 1983 he was with NCR Corporation as a system engineer and project leader. From 1983 to 1985 he was an assistant professor in the Department of Computing and Information Science, University of Guelph, Ontario. He has been a faculty member in the Department of Systems Design Engineering since 1985. He is at present an associate professor and an associate director of the Pattern Analysis and Machine Intelligence Laboratory. Dr. Kamel is a member of the Institute of Electrical and Electronics Engineers, the Association of Computing Machinery, and the American Association of Artificial Intelligence.

Helen C. Shen *PAMI Group, Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.* Dr. Shen is an associate professor in the Department of Systems Design Engineering. She holds a B.Math degree from the University of Waterloo and an M.Sc. in computer science from the University of Toronto. She obtained her Ph.D. from Waterloo University in 1982. Dr. Shen's areas of research include texture analysis (both monochrome and color); computer vision; sensory data integration in an autonomous workcell; autonomous workcell configurations; error detection, identification, and recovery in robotics; and parallel algorithm design.

Andrew K. C. Wong *PAMI Group, Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.* Dr. Wong received his Ph.D. from Carnegie Mellon University in 1968. For several years he taught at Carnegie Mellon. He is currently a professor of Systems Design Engineering and director of the Pattern Analysis and Machine Intelligence Laboratory at the University of Waterloo and an honorable professor at the University of Hull, United Kingdom. Dr. Wong has authored and coauthored chapters and sections in a number of books on engineering and computer science and has published many articles in scientific journals and conference proceedings. He is the winner of the FCCP 1991 Award of Merit.

Radu I. Campeanu *Glendon College, York University, Computer Science Department, 2275 Bayview Avenue, Toronto, Ontario M4N 3M6, Canada.* Dr. Campeanu is currently an associate professor in computer science at Glendon College. He received an M.Sc. in physics from the University of Cluj, Romania, in 1972 and a Ph.D. in computational atomic physics from the University College London, United Kingdom, in 1977. Between 1977 and 1985 he taught in the Department of Physics at the University of Cluj. In 1986, Dr. Campeanu joined the IBM Canada Laboratory, where he worked in the Image Systems Centre and in the Centre for Advanced Studies. He moved to his current position in 1991.

Reprint Order No. G321-5515.