Face Recognition from Profiles Using Morphological Signature Transform

Zdravko Liposcak and Sven Loncaric

Zdravko Liposcak, B.Sc.
MOI RH, Rimska 19, Sisak, Croatia, Tel: +385 44 56 01 30,
E-mail: zliposcak@yahoo.com

Sven Loncaric, Ph.D.
University of Zagreb, Faculty of Electrical Engineering and Computing,
Department of Electronic Systems and Information Processing, Unska 3, Zagreb, Croatia
Tel: +385 1 612 9891, +385 1 612 9911, E-mail: sven.loncaric@fer.hr

Abstract

A new method for face recognition using profile gray-level images based on the morphological signature transform (MST) is presented in this paper. Gray-level image of profile is thresholded to produce a binary, black and white image. The method is based on the successive morphological erosions of the input profile at different resolutions by primary and rotated structuring elements. The areas of successively eroded images are computed for each structuring element at each pyramid level. The obtained set of numbers is arranged into vectors, ordered, and used as a face descriptor. Obtained recognition rates are 90% for the distorted images (synthetic data) and 76.6 % for the unknown profile image of the stored face.

key words: face recognition, morphological signature transform, shape representation.

1. Introduction

Since the early 1990's Face Recognition Technology (FRT) become an active research area.

A general statement of the problem of face recognition can be formulated as follows: Given still or video images of a scene, identify one or more persons in the scene using a stored database of faces [2]. The solution of the problem involves segmentation of faces from cluttered scenes, extraction of features from face region, identification, and matching.

Face recognition problems and techniques can be separate in two broadly groups: dynamic (video) and static (no video) matching. Dynamic matching will be used when a video sequence is available. The video images tend to be of low quality, the background is very cluttered and often is more than one face present in the picture. However, since a video sequence is available, one could use motion as a strong cue for segmenting faces of moving persons. Static matching used images with typically reasonably controlled illumination, background, resolution, and distance between camera (or 3D scanner) and the person. Some of the images that arise in this group can be acquired from a video camera. Machine recognition of faces has several applications, ranging from static matching of controlled photographs as in mug shots matching and credit card verification to surveillance video images. Mug shots matching is the most common application in static matching group. Typically, in mug shots photographs one frontal and one or more side views of a person's face are taken. Profile images provide a detailed structure of the face that is not seen in frontal images.

In this work we present a method for face representing and recognition from profile face images. For that purpose we use the extracted area of profile for constructing the shape that will be represent the face. First of all, we give a brief description of the Morphological Signature Transform (MST), a new approach for morphological shape representation [1], [8], from profile images using the MST. The short algorithm with some basics steps is also included. Finally, results and the conclusions are given.

2. Morphological signature transform (MST) for shape description

Loncaric and Dhawan [1], [8] developed a method for shape description called Morphological Signature Transform (MST). The MST method for shape description utilizes multi-resolution morphological image processing by multiple structuring elements (SEs). The MST shape representation method is based on the decomposition of a *complex* shape to multiple *simple* signature shapes. The idea of this approach is to process decomposed, multiple shapes instead of the original shape which is extracted through the property decomposition process. A brief description of the MST is given below.

For generality, let W be a set, which contains shapes to be described. Some common examples for W are: $W = R^2$ (planar shapes) or $W=R^3$ (three-dimensional shapes), where R denotes the set of real numbers. The multiplication of a set by a real number is defined as

$$rX = \bigcup_{x \in X} \{rX\} \tag{1}$$

where $r \in R$ and $X \subset W$. Another notation that we use in this paper is: $S^n = S \oplus S \oplus ... \oplus S$, where there are n summands in the Minkowski addition on the right side of the equation. Minkovski set addition defined as

$$X \oplus Y = \bigcup_{y \in Y} X_y \tag{2}$$

Note that generally $S^n \neq nS$, but in the special case when S is a convex set it holds that $S^n = nS$.

The MST of shape X with respect to structuring element S is defined as:

$$X_s(r,n) = M(rX, S^n) \tag{3}$$

where $r \in R$, $n \in N$ and M is binary morphological operator, i.e. $M: P \times P \rightarrow P$, where P is the set of all subsets (i.e. the partitive set) of W. In other words M takes two sets from P as arguments and produces a third set as result. Some common examples for binary operator M are morphological erosion, dilation, opening, and closing. To compute the MST a morphological operator M must be selected. The morphological erosion has been used in the work presented in this paper. For a given set X, and structuring element S a family of sets $X_S(r,n)$ is generated by varying the parameters r and n. The parameter r is a scale parameter. The parameter n plays the role in changing the size and the shape of the structuring element S. Generated shapes $X_S(r,n)$ are called the signature shapes of X with respect to S. Signature shapes $X_S(r,n)$ can be used to characterize the original shape X.

The idea of MST approach is to process decomposed, multiple shapes instead of processing the original shape. In such a way, a problem of extracting complex shape properties from a single object is replaced with a simpler problem of extracting a simple shape property from derived multiple signature shapes. The algorithm uses area as the basic shape descriptor instead of residue, which is more prone to noise. The MST-based shape description method has the following properties: 1) translation invariance; 2) size invariance; 3) rotation invariance.

3. Face representation and recognition using MST

The input in our face processing and analysis system is a gray-scale image of a scene containing the side view of a human face. In this work, we used a collection of profile face images University of Bern[1] for learning and recognition of faces. This image database contains profile views of 30 people. In first step images was tresholded to produce a binary, black and white image, the white corresponding to the face region. A pre-processing step then extracts the front portion of the silhouette that bounds the face image. This is to avoid variations in the profile due to changes in hairline [3].

Shape that will be representing the profile face was created from the extracted part of the silhouette. First of all, we extracted two fiducial points A and B [4], [5], [6]. This two points are eye point and chin point, with co-ordinates $A...(x_1,y_1)$ and $B...(x_2,y_2)$. Basic dimension for the constructing of shape is distance d ($2*d$ is a distance between point A and point B);

$$d = \left\{ \left[(x_2 - x_1)^2 + (y_2 - y_1)^2 \right]^{0.5} \right\} / 2 \tag{4}$$

Fig.1.a. showing the process of extracting front part of profile face and the constructing of face shape. The obtained face shape is linearly scaled in both dimensions in such a way that the shape area is equal to some constant value. This step assures size invariance of the method.

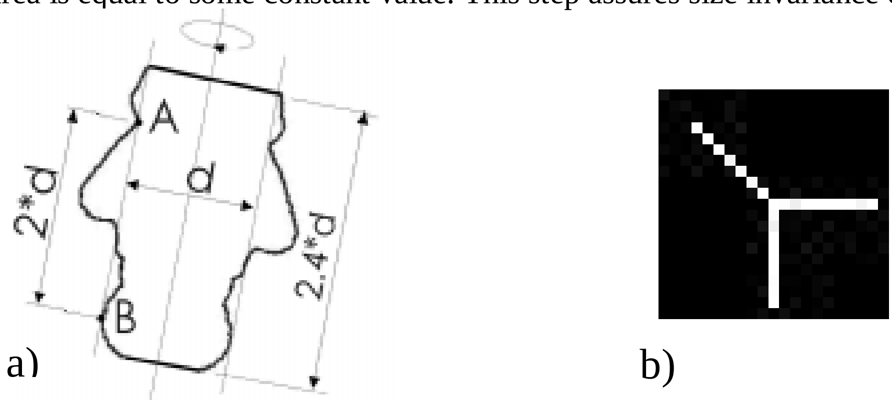

Fig. 1. a) Process of constructing the profile face shape b) Structuring element (Y)

A shape property that is used to describe signature shapes in this work is the area. It is both simple and easy to compute- for binary images it is just a count of pixels with values equal to one. The signature shapes are created by means of the MST based on the morphological erosion. In that case, the signature shape equation (3) becomes

$$Z(r, n) = \mathcal{E}(rX, S^n) \tag{5}$$

and the area $A(Z(r,n))$ of the signature image $Z(r,n)$ is computed for some $r \in R$ and $n \in N$. In addition, multiple rotated versions of the structuring element S are used in equation (5), as explained below, to ensure additional shape information as well as the rotation invariance of the method. In this work we use non-convex structuring elements (see Fig.1.b.) which were created to sense both detailed and coarse structure of the profile image.In the case of non-convex structuring elements, multiple resolutions and successive erosions provide additional information about the object shape.

Multiresolution and successive erosions are redundant in the case of convex structuring elements. Rotational invariance is achieved through rotation of one fixed structuring element. The structuring elements used in this paper are created *empirically* to detect both the detailed and coarse structure of the object. A method for near-optimal shape matching using MST was developed in reference [1]. It is based on a genetic algorithm for selection of a near-optimal structuring element for MST shape description. We use *standard* "Y" structuring element (who is not selected by using proposal optimization method [1]).

The structuring element must have a 360^0 period of rotational symmetry; such a requirement comes from the nature of the MST method. The MST method could also utilize structuring elements with some other period of rotational symmetry, although the 360^0 case provides the maximum amount of information, given some fixed rotational resolution. From the initially chosen structuring element we derive a set of N structuring elements by rotating initial one by the angle α;

$$\alpha = i\frac{360}{N}, \qquad i = 1,\ldots,N-1 \tag{6}$$

In such a way, we are able to achieve rotational invariance of the method. The basic steps of the algorithm are as follows:

1. The input graylevel image is tresholded to produce a binary, black and white image X_f, the white corresponding to face region.
2. Construct the profile face shape using two fiducial points, eye point and chin point. Extracted shape X representing the profile face image X_f .
3. The binary image X is linearly scaled in both dimensions in such a way that the shape area is equal to some constant value; A(X) = cons. This step assures size invariance of the method. Call the scaled image Y_i^0, where the subscript *i=0 to N-1* denotes the rotated version of the scaled structuring element while the superscript denotes the pyramid level.
4. The four level multiresolution pyramid is created by subsampling by factor of two from the normalized version of the original image obtained in the previous step (Y_i^1, Y_i^2, Y_i^3, ...).
5. The *N* rotated versions of the structuring element are created from the original one.
6. For each *j* multiresolution pyramid level do *(j = 0, ... , R-1)*
   (a) For each *i* rotated version of the structuring element do *(i = 0, ... , N-1)*
      i. Compute successive erosions of the input image by a given structuring element and at a given pyramid level and store the areas of successively eroded images in a vector. $A_i^j(n) \leftarrow A(\mathcal{E}(Y_i^j, S^n))$, where $n = 1, \ldots, L$ (constant L is the number of successive erosions computed).
      ii. $a_i \leftarrow \left[A_1^j A_2^j \ldots A_L^j\right]^T$, where the superscript T denotes the vector transposition. Form a vector $a_i \in R^L$ from L successive areas computed for the *j*th pyramid level.
   (b) $b_j^T = \left[a_0^T a_1^T \ldots a_{N-1}^T\right]$ Form a vector $b_j \in R^{LN}$ from N vectors a_i computed for pyramid levels of the *i*th rotated structuring element.
   (c) Circularly shift the vectors (to achieve rotation invariance) so that the largest one, call it b_g , with respect to the Euclidean norm, comes to the zeroth position. $c_j \leftarrow b_{(j+g) \bmod N}$
   (d) Take one or more vectors from the shifted set as a representative for this particular pyramid level.
7. Take the vector(s) resulting from each pyramid level and store them in one final shape descriptor vector $d_x \leftarrow \left[c_0^T c_1^T \ldots c_{R-1}^T\right]^T$. Form the final morphological face descriptor $d_x \in R^{LNR}$ of the face *X* by concatenating R descriptor vectors for each pyramid level.

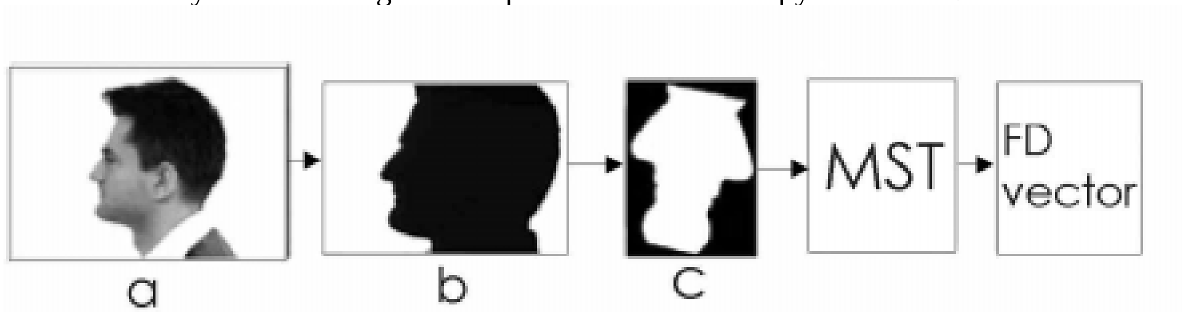

Fig. 3. The basic steps of the MST algorithm; a) input graylevel image, b) binary image, c) constructed shape and face description vector (FDV)

8. To achieve more realistic distance measure, vectors from different pyramid levels may be weighted, yielding the final morphological face descriptor vector.

The *profile face vector* obtained in the final step of the algorithm is the final result of the algorithm and is used as a *morphological face descriptor*. As described above, such a shape descriptor has all the desired invariance properties- invariance to rotation, size, and translation.

4. Results and discussion

Using the MST shape descriptor algorithm we obtained the vector corresponding to each input profile face image. The vector is called the *morphological face descriptor*. The dimensions of the morphological face descriptor depends on the number of multiresolution pyramid levels, the number of the rotated versions of the original structuring element, and the number of successive erosions computed. Theoretically, the more rotated version we have the closer to rotation invariance we get. In practice, however, there exists a maximum number of rotations (maximum resolution) after which additional rotations do not improve accuracy, due to finite precision and errors of the rotation operation.

The study of similarities and differences between faces now reduces to the study of similarities and differences between morphological face descriptor vectors. If we have two face descriptors x, and y, of dimension k, then we can take a distance between them as a measure of the similarity of the faces they represent. The smaller the distance, the more similar the faces. The Euclidean distance had been used in this work

$$d_2(x,y)=\left(\sum_{i=1}^{k}(x_i-y_i)^2\right)^{1/2} \qquad (7)$$

For training and testing set we use a profile face images Database University of Bern which contains profile views of the 30 people. For each person they took a five graylevel images with variations of the head position, the size and the contrast - 1, 2, 3 big profiles with high contrast, looks like binarized, 4, 5 small profiles with normal gray levels. Pictures are with controlled uniform background and without background clutter. Size of this images is 342×512 pixels. Structuring element of size 17×17 are shown in Fig. 2. Four pyramid levels were used (R = 4) to represent the input profile image. The structuring element were rotated into 36 different positions from 0^0 to 350^0 (N = 36). The successive area sequence length was 4 elements for the each rotation (L = 4). This gave a total descriptor length of 144 elements for one pyramid level. The total profile descriptor size for all 4 pyramids level was then equal to 576. In the [8], [1] authors propose different and more precisely approach. They rotating input images instead of rotating structuring element. Structuring elements are typically of small size (17 by 17 is used in this work). Rotating an image of such a small size by a small angle typically gives an inaccurate result. They rotated images into 64 different position from 0^0 to 360^0, and used two structuring elements resulting in a 1920- dimensional shape descriptor. We performed two kinds of sequence weighting proposed in [8]. The successive area sequence is weighted for some particular pyramid level and rotation position by the empirically derived formula

$$w_p(n)=\begin{cases}1/\left(1-\frac{n}{1.5(p-1)}\right) & 1\le n< p\\ 0 & p\le n\le L\end{cases} \quad , \; A_i^j(n)\leftarrow A_i^j(n)w_p(n), \qquad n=1,\ldots,L \qquad (8)$$

where p is the index of the last non-zero element in the sequence. The use of this weighting expression enhances exactly the middle and most important range of the sequence.

The pyramid level weighting is done with respect to pyramid level. Since area decreases quadratically as resolution decreases by factor of two this weighting is performed by multiplying all areas in pyramid level *j* by 4^j, j=1,...,R-1, where R is the total number of pyramid levels.

We are performed two type of experiments.

The first experiment consisted in shape matching between distorted and original images. A set of 30 profile images, one from each person is used as original images. Another set of 30 images was created from the original images, by shape distortion. Shape distortion includes small changes of size, orientation (rotation), and addition of noise to the profile boundary. A recognition rate of 90.0% between original and distorted images was obtained.

In second experiment we use a 4 profile images each of 30 people in training set. The fifth profile face image each of 30 person we use in testing set. Algorithm is successively recognized 23 of 30 persons. Seven faces were recognized incorrectly. Obtained profile face images recognition rate is 76.6%.

5. Conclusion

In this work we use simpler version of novel shape analysis technique presented by Loncaric and Dhawan [8], [1], [9] called Morphological Signature Transform (MST) for purpose of recognition face from profile face images. The proposed algorithm is focused on extraction of features from profile region and recognition of faces. Obtained recognition rates are 90% for the distorted images (profile distortion includes small changes of size, rotation, and addition of noise to the profile boundary) and 76.6 % for the unknown profile image of the stored face. Recognition rates can be possibly improved in several ways:

i. using rotated face image instead of rotated structuring element
ii. using multiple structuring elements and more training pictures
iii. using more levels of successive erosions and multiresolution pyramid levels
iv. using optimal structuring element for given set of profile face images [1]

There exists a maximum numbers of rotations, structuring elements, levels of successive erosions and multiresolution levels after which accuracy can't be improve, due to finite precision.

References

1. Loncaric, S. and Dhawan A. P. (1995), "Near-optimal MST-based shape description using genetic algorithm," Patt. Recog., vol. 28, pp. 571-579.
2. Chellappa, R. Wilson, C. L. and Sirohey, S. (1995), "Human and Machine Recognition of Faces: A survey," Proc. of the IEEE, vol. 83, No. 5, pp. 705-740.
3. Kaufman, G. J. Jr., and Breeding, K. J. (1976), " The automatic recognition of human faces from profile silhouettes," IEEE Trans. Syst., Man and Cybern, vol. SMC-6, pp. 113-121.
4. Harmon, L. and Hunt, W. (1977)," Automatic recognition of human face profiles," Computer Graphic and Image Process., vol. 6, pp. 135-156.
5. Harmon, L. Kuo, S. Ramig, P. and Raudkivi, U. (1978), "Identification of human face profiles by computer," Patt. Recog., vol. 10, pp. 301-312.
6. Wu, C. and Huang, J. (1990), "Human face profile recognition by computer," Patt. Recog., vol. 23, pp. 255-259.
7. Aibara, T. Ohue, K. and Matsuoka, Y. (1991), "Human face recognition of P-type Fourier descriptors," in SPIE Proc., vol. 1606: Visual Commun. and Image Process, pp. 198-203.
8. Loncaric, S. and Dhawan, A. P. (1993), "A morphological signature transform for shape description", Patt. Recog. vol. 26, pp. 1029-1037.
9. Loncaric, S. (1998), "A survey of shape analysis techniques", Patt. Recog. vol. 31, pp.983-1001.