Human Recognition Based on Face Profiles in Video

Xiaoli Zhou and Bir Bhanu
Center for Research in Intelligent Systems
University of California, Riverside
Riverside CA 92521
{xzhou, bhanu } @vislab.ucr.edu

Abstract

Face profile is an important aspect of face recognition and it provides a complementary structure of the face that is seen in the non-frontal view. In the past, several methods have been proposed to recognize face profiles in still images. However, face profile images that are captured at a distance by surveillance cameras usually are video sequences that have a low resolution. It is difficult to extract accurate face profile directly from a low-resolution video frame, which does not have many pixels on the face profile. The emphasis of this paper is to introduce a practical approach for human recognition by using high-resolution face profile images constructed from the low-resolution videos. We use both the spatial and temporal information present in a number of adjacent low-resolution frames of a video sequence to construct high-resolution face profile images. As the quality of high-resolution images relies on the correctness of image alignment between consecutive frames, an elastic registration algorithm is used for face profile image alignment. A match statistic is designed to detect and discard poorly aligned images which may degrade the quality of the high-resolution face profile image. After obtaining high-resolution face profile images, we use a dynamic time warping method for face profile recognition. A number of dynamic video sequences are tested to demonstrate the applicability and reliability of our method.

1 Introduction

During the last decades, many methods have been proposed to recognize face profiles in still images. However, human recognition based on face profiles in video has achieved little attention in spite of the fact that in surveillance scenarios where human is walking at a distance, generally the face profile or the side face, not the frontal face will be available. Whenever we use a video camera in order to record a dynamic scene, we encounter many challenges. Since a video camera is characterized by finite spatial and temporal resolutions, objects in the scene moving with a high velocity compared to the frame rate are blurred, and their shape is distorted. Moreover, non-ideal detectors and poor optics introduce spatial blur, which affects the sharpness of the images [1]. Consequently, the face profiles of a moving person in video are blurred in some way and not as sharp as that in a still image. Also, when the video is taken at a distance, a face profile has a very limited resolution. It is difficult to get accurate information from such a face profile, which degrades the performance of a face profile recognition system and limits the application of face profile in surveillance video.

This work aims to introduce a practical approach for human recognition by using high-resolution face profile images that are constructed from the low-resolution videos. We use both the spatial and temporal information present in a number of adjacent low-resolution frames of a video sequence to construct high-resolution face profile images. This relies on the fact that temporally adjacent frames in a video sequence, in which a human is walking with a side view to the camera, contain slightly different, but unique, information for face profile. The quality of reconstructed face profile images is much better than those directly obtained from the videos. Since the quality of high-resolution images relies on the correctness of image alignment between consecutive frames, an elastic registration algorithm is used for face profile image alignment. A match statistic is designed to detect and discard poorly aligned images. After obtaining high-resolution face profile images, we use a dynamic time warping method for face profile recognition.

2 Technical Approach

The overall technical approach is shown in Figure 1. It consists of the following key steps: side face image extraction from video, registration of side face images, high-resolution side face images construction, and face profile extraction and recognition.

2.1 Automatic Side Face Image Extraction

Correlation is useful for feature detection. In our work, a correlation-based method is used for side face image detection and extraction. For an image of size MxN and a template window of size JxK, we compute the normalized cross correlation (NCC) between them,

$$r(u,v) = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K} f_1(j,k) f_2(j-u,k-v)}{[\sum_{j=1}^{J}\sum_{k=1}^{K} f_1^2(j,k)]^{1/2}} \times \frac{1}{[\sum_{j=1}^{J}\sum_{k=1}^{K} f_2^2(j-u,k-v)]^{1/2}} \quad (1)$$

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 **IEEE**

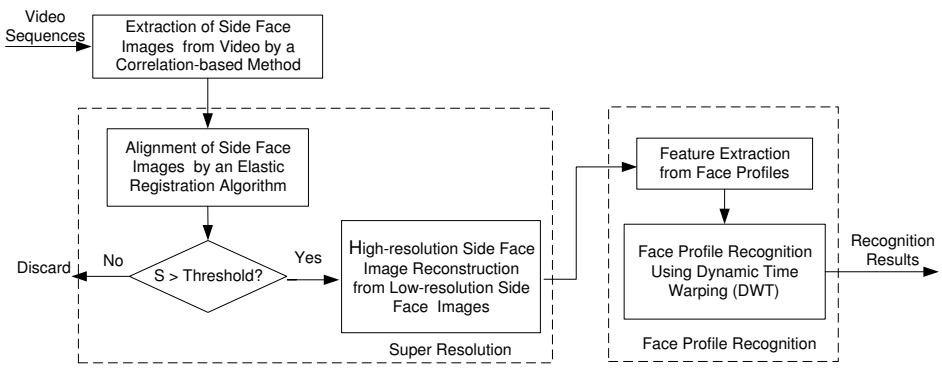

Figure 1: Face profile recognition in video.

where (j,k) are indices in a JxK point window area. For an image that contains a target feature, the NCC forms local maxima in candidate positions. To increase the speed, before applying the NCC, we first detect the region where the target is most likely to be present. Then we calculate the NCC between the template and the defined region, instead of the whole frame. The steps of our method are as follows:

1. Extract a small image, called the template, containing just the feature to be detected from a video frame.

2. Subtract the background frame from the source frame and detect the region where motions (changes) occur.

3. Calculate the NCC between the template and the region defined in the previous step.

4. Measure candidate region locations and extract the region, where the NCC get the local maxima, as the side face image.

In our work, the template is a small image containing the side view of a person's head. Some results are shown in Figure 2.

2.2 Side Face Image Registration

The quality of high-resolution images relies on the correctness of image alignment between consecutive frames. Thus, motion estimates must be computed to determine pixel displacements between frames. In our work, an elastic registration algorithm [2] is used for motion estimation of face profiles. Also, in order to obtain the optimal high-resolution image, we design a match statistic to detect and discard images that are poorly aligned and may degrade the output quality. Hence, the quality of high-resolution images can be improved by rejecting such errors.

2.2.1 Local affine model with Intensity Variations

Denote $f(x, y, t)$ and $f(\hat{x}, \hat{y}, t-1)$ as the source and target images, respectively. The model first assumes that the image intensities between images are conserved, and that the motion between images can be modeled locally by an affine transform:

$$f(x, y, t) = f(m_1 x + m_2 y + m_5, m_3 x + m_4 y + m_6, t-1) \quad (2)$$

where m_1,m_2,m_3,m_4 are the linear affine parameters, and m_5,m_6 are the translation parameters. The previous assumption that the image intensities between the source and target are unchanged is likely to fail under a number of circumstances. To account for intensity variations, an explicit change of local contrast and brightness is incorporate into the affine model. Specifically, the initial model takes the form:

$$m_7 f(x, y, t) + m_8 = f(m_1 x + m_2 y + m_5, m_3 x + m_4 y + m_6, t-1) \quad (3)$$

where m_7 and m_8 are two new (also spatially varying) parameters that embody a change in contrast and brightness, respectively. In order to estimate these parameters, the following quadratic error function is to be minimized:

$$E(\vec{m}) = \sum_{x,y \in \Omega} [m_7 f(x, y, t) + m_8 - f(m_1 x + m_2 y + m_5, m_3 x + m_4 y + m_6, t-1)]^2 \quad (4)$$

where $\vec{m} = (m_1 ... m_8)^T$, and Ω denotes a small spatial neighborhood. Since this error function is nonlinear in its unknowns, it cannot be minimized analytically. To simplify the minimization, this error function is approximated by using a first-order truncated Taylor series expansion. It now takes the form below.

$$E(\vec{m}) = \sum_{x,y \in \Omega} [k - \vec{c}^T \vec{m}]^2 \quad (5)$$

where the scalar k and vector $\vec{c}$ are given as:

$$k = f_t + x f_x + y f_y$$
$$\vec{c} = (x f_x \quad y f_x \quad x f_y \quad y f_y \quad f_x \quad f_y)^T$$

where $f_x(\cdot), f_y(\cdot), f_t(\cdot)$ are the spatial/temporal derivatives of $f(\cdot)$. Minimization of this error function is accomplished

(a) (b) (c) (d)

Figure 2: Low-resolution side face images extracted from videos using the normalized cross correlation.

(a)S=0.9411 (b)S=0.9314

(c)S=0.9420 (d)S=0.9546

Figure 3: The reference image (left) and the aligned image (right) with the match statistic S.

by differentiating $E(\vec{m})$, setting the result equal to zero and solving for $\vec{m}$. The solution is,

$$\vec{m} = [\sum_{x,y\in\Omega} \vec{c}\vec{c}^T]^{-1}[\sum_{x,y\in\Omega} \vec{c}k] \tag{6}$$

Intensity variations are typically a significant source of error in differential motion estimation. The addition of the contrast and brightness terms allows us to accurately register images in the presence of local intensity variations. The model is given another assumption that the model parameters $\vec{m}$ vary smoothly across space. A smoothness constraint on the contrast/brightness parameters has the added benefit of avoiding a degenerate solution where a pure brightness modulation is used to describe the mapping between images.

To begin, the error function is augmented as follows:

$$E(\vec{m}) = E_b(\vec{m}) + E_s(\vec{m}) \tag{7}$$

where $E_b(\vec{m})$ is defined without the summation:

$$E_b(\vec{m}) = \sum_{x,y\in\Omega} [k - \vec{c}^T\vec{m}]^2 \tag{8}$$

with k and $\vec{c}$ as before. The new quadratic error term $E_s(\vec{m})$ embodies the smoothness constraint:

$$E_s(\vec{m}) = \sum_{i=1}^{8} \lambda_i[(\frac{\partial m_i}{\partial x})^2 + (\frac{\partial m_i}{\partial y})^2] \tag{9}$$

where λ_i is a positive constant that controls the relative weight given to the smoothness constraint on parameters. This error function is again minimized by differentiating with respect to the model parameters, setting the result equal to zero and solving $\frac{dE(\vec{m})}{d\vec{m}} = \frac{dE_b(\vec{m})}{d\vec{m}} + \frac{dE_s(\vec{m})}{d\vec{m}} = 0$. Instead $\vec{m}$ is expressed in the following form:

$$\vec{m}^{(j+1)} = (\vec{c}\vec{c}^T + L)^{-1}(\vec{c}k + L\overline{\vec{m}}^{(j)}) \tag{10}$$

where $\overline{\vec{m}}$ is the component-wise average of $\vec{m}$ over a small spatial neighborhood, and L is an 8x8 diagonal matrix with diagonal elements λ_i, and zero off the diagonal. An iterative scheme is employed to solve for $\vec{m}$. On each iteration j, $\overline{\vec{m}}^j$ is estimated from the current $\vec{m}^{(j)}$. The initial estimate $\vec{m}^{(0)}$ is estimated from Equation 6.

2.2.2 Match Statistics

A match statistic indicates how well a transformed image aligns with the reference image. It is used to select or reject transformations during alignment. If the size of the reference image is MxN, the mean square error E between the aligned image and the reference image is $\sum_{x=1}^{M}\sum_{y=1}^{N}[f(x,y,t) - f(m_1x+m_2y+m_5, m_3x+m_4y+m_6, t-1)]^2/MN$. So the match statistic of the aligned images is defined as

$$S = 1 - \frac{E}{[\sum_{x=1}^{M}\sum_{y=1}^{N} f^2(x,y,t)]/MN} \tag{11}$$

If the value is close to 100%, then the image is well aligned. A very low value indicates misalignment. However, even images that are very well aligned typically do not achieve 100%. For improving image quality, the method

works most effectively when match values are lower than 100%.

A match threshold can be specified, which means no images with the final match statistic falling below this threshold will contribute to the final output. A perfect match is 100%. Zero means any image can be accepted. Some alignment results are shown in Figure 3.

2.3 High-Resolution Side Face Image Construction

Multiframe resolution enhancement seeks to construct a single high-resolution image from several low-resolution images. These images must be of the same object, taken from slightly different angles, but not so much as to change the overall appearance of the object in the image. The idea of super-resolution was first introduced in 1984 by Tsai and Huang [3] for multiframe image restoration of band-limited signals. In the last two decades, different mathematical approaches have been developed. All of them seek to address the question of how to combine irredundant image information in multiple frames. A good overview of existing algorithms is given by Borman and Stevenson [4] and Park et al. [5]. In this paper, we use an iterative method proposed by Irani and Peleg [6][7] to construct high-resolution side face images from low-resolution side face images. These low-resolution side face images have been aligned using the elastic registration algorithm in Section 2.2.

2.3.1 The Imaging Model

The imaging process, yielding the observed image sequence g_k, is modeled by:

$$g_k(m,n) = \sigma_k(h(T_k(f(x,y))) + \eta_k(x,y)) \tag{12}$$

where

1. g_k is the sensed image of the tracked object in the k_{th} frame.
2. f is a high resolution image of the tracked object in a desired reconstruction view. Finding f is the objective of the super-resolution algorithm.
3. T_k is the 2-D geometric transformation form f to g_k, determined by the computed 2-D motion parameters of the tracked object in the image plane (not including the decrease in sampling rate between f and g_k). T_k is assumed to be invertible.
4. h is a blurring operator, determined by the Point Spread Function of the sensor (PSF). When lacking knowledge of the sensor's properties, it is assumed to be a Gaussian.
5. η_k is an additive noise term.
6. σ_k is a downsampling operator which digitizes and decimates the image into pixels and quantizes the resulting pixels values.

The receptive field (in f) of a detector whose output is the pixel $g_k(m,n)$ is uniquely defined by its center (x,y) and its shape. The shape is determined by the region of the blurring operator h, and by the inverse geometric transformation T_k^{-1}. Similarly, the center (x,y) is obtained by $T_k^{-1}((m,n))$. An attempt is made to construct a higher resolution image $\hat{f}$, which approximates f as accurately as possible, and surpasses the visual quality of the observed images in $\{g_k\}$

2.3.2 The Super Resolution Algorithm

The algorithm for creating higher resolution images is iterative. Starting with an initial guess $f^{(0)}$ for the high resolution image, the imaging process is simulated to obtain a set of low resolution images $\{g_k^{(0)}\}_{k=1}^K$ corresponding to the observed input images $\{g_k\}_{k=1}^K$. If $f^{(0)}$ were the correct high resolution image, then the simulated images $\{g_k^{(0)}\}_{k=1}^K$ should be identical to the observed image $\{g_k\}_{k=1}^K$. The difference images $\{g_k - g_k^{(0)}\}_{k=1}^K$ are used to improve the initial guess by "backprojecting" each value in the difference images onto its receptive field in $f^{(0)}$, yielding an improved high resolution image $f^{(1)}$. This process is repeated iteratively to minimize the error function:

$$e^{(n)} = \sqrt{\frac{1}{K}\sum_{k=1}^{K} \|g_k - g_k^{(0)}\|_2^2} \tag{13}$$

The imaging process of g_k at the n_{th}iteration is simulated by:

$$g_k^{(n)} = (T_k(f^{(n)}) * h) \downarrow s \tag{14}$$

where $\downarrow s$ denotes a downsampling operator by a factor s, and * is the convolution operator. The iterative update scheme of the high resolution image is expressed by:

$$f^{(n+1)} = f^{(n)} + \frac{1}{K}\sum_{k=1}^{K} T_k^{-1}(((g_k - g_k^{(n)}) \uparrow s) * p) \tag{15}$$

where K is the number of low resolution images. $\uparrow s$ is an upsampling operator by a factor s, and p is a "backprojection" kernel, determined by h and T_k as explained below. The averaging process reduces additive noise. The algorithm is numerically similar to common iterative methods for solving sets of linear equations, and therefore has similar properties, such as rapid convergence.

2.4 Face Profile Recognition

For face profile recognition, we use a curvature-based matching approach [8], which does not focus on all fiducial point extraction and the determination of relationship among these fiducial points like most of current algorithms do, but attempt to use as much information as a profile possesses.

2.4.1 Face Profile Extraction

The outline of a profile is treated as a function, which consists of a set of points $T(u) = (x(u), y(u))$, including fiducial points like nasion, pronasale, chin and throat. The following is the procedure, which is used to extract the face profile contour from the side view images.

1. Apply canny edge detector to the high-resolution side face image and obtain a binary image.
2. Extract the outline curve of the front of the silhouette (the profile line) as the face profile contour by extracting the leftmost point different from background.
3. The 2D curve can be regarded as 1D function $T(u)$, where x is a row index and y is a column index of a pixel inside a profile line.

2.4.2 Curvature Estimation

In our method, the absolute values of curvature are computed on face profile contours. Gaussian filtering is applied before curvature is estimated. It aims to eliminate the spatial quantization noise introduced during the digitization process, as well as other types of high frequency noise. The Gaussian convolution of $x(u)$ and $y(u)$ depends both on u, the signal's variable, and σ, the Gaussian's standard deviation. The Gaussian kernel is given by

$$g(x, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}} \tag{16}$$

Curvature on $T_\sigma(u) = (X(u, \sigma), Y(u, \sigma))$ is given as:

$$\kappa(\mu, \sigma) = \frac{X_\mu(\mu, \sigma) Y_{\mu\mu}(\mu, \sigma) - X_{\mu\mu}(\mu, \sigma) Y_\mu(\mu, \sigma)}{(X_\mu(\mu, \sigma)^2 + Y_\mu(\mu, \sigma)^2)^{1.5}} \tag{17}$$

where the first and second derivatives of X and Y is represented as X_u, X_{uu}, Y_u and Y_{uu}, respectively

Using the curvature value, the fiducial points, including the nasion and throat can be reliably extracted using a fast and simple method in [8].

2.4.3 Profile Matching Using Dynamic Time Warping

A dynamic time warping method is applied to compute the similarity of two face profiles from nasion to throat. The absolute values of curvature are used to represent the shapes of face profile contours. The dynamic time warping is an algorithm to calculate the optimal score and to find the optimal alignment between two strings. This method is a more robust distance measure for time series than Euclidean distance. For two strings $s[1, ..., n]$ and $t = [1, ..., m]$, we compute $D(i, j)$ for entire sequences, where i ranges from 1 to m and j ranges from 1 to n. $D(i, j)$ is defined as:

$$\begin{aligned} D(i,j) = min\{&D[i-1, j-1] + d(s[j], t[i]), \\ &D[i-1, j] + gap, \\ &D[i, j-1] + gap\} \end{aligned} \tag{18}$$

Here, $d(s[j], t[i])$ represents the similarity between two points on face profiles. Since the face profile is represented by the curvature of the profile, $d(s[j], t[i])$ is calculated by Euclidean distance:

$$d(s[j], t[i]) = ||s[j] - t[i]|| \tag{19}$$

The penalty is defined for both horizontal and vertical gaps. It should be small and yet exist just to control non-diagonal moves. Generally, the penalties should be set to less than 1/10th the maximum of the $d(s[j], t[i])$. In our method, we use the same constant penalty for both horizontal and vertical gaps. The final score $D(m, n)$ is the best score for the alignment. The smaller $D(m, n)$ is, the higher the similarity.

Figure 4 shows the face profile extracted from a high-resolution image and the absolute values of curvature. Figure 5 gives an example of dynamic time warping of two face profiles from the same person. From the similarity matrix in Figure 5, we can see a light stripe (high similarity values) approximately down the leading diagonal. From the dynamic programming matrix in Figure 5, we can see the lowest-cost path between the opposite corners visibly follows the light stripe, which overlay the path on the similarity matrix. The least cost is the value in the bottom-right corner of the dynamic programming matrix. This is the value we would use to compare between different templates when we are doing classification.

3 Experimental Results

The data is obtained by Sony DCR-VX1000 digital video camera recorder. We collect 28 video sequences of 14 people who are walking outside and presenting a side view to the camera. The distance between people and the video camera is about 7 feet. Figure 6 shows some examples of our data. Each of the persons has two sequences. For 4 of the subjects, the data was collected on two separate days and about 1 months apart. Lengths of the sequences range from 80 to 140, with an average of 100 frames, at 30 frames per second. The shutter speed is 1/60 and the resolution of each frame is 720x480.

In our system, we reconstruct a high-resolution face profile image from 11 low resolution face profile images. These low-resolution face profile images are localized and extracted from 11 adjacent video frames by the correlation-base method. After aligning the other 10 low-resolution face profile images with the reference image using the elastic registration algorithm, we apply the super resolution algorithm to construct a high-resolution face profile image. The low-resolution reference image is typically the face profile image corresponding to the middle one among 11 consecutive video frames in the sequence. In our work, a 70x70 pixel reference image is first expanded to a 140x140 pixel image by nearest neighbor or bilinear interpolation method. The upsampled reference image will simply be a blurred version of the original, but at the desired higher resolution. The goal now is to use the information from the rest

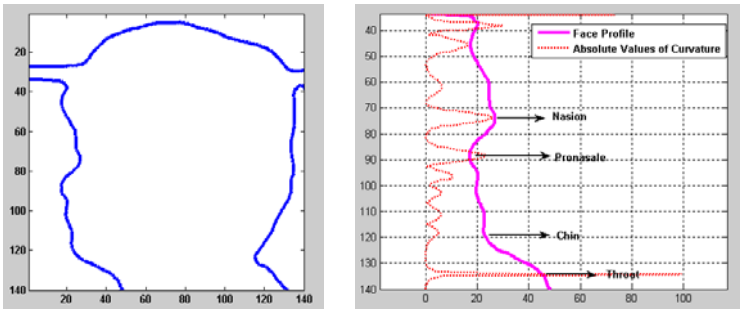

Figure 4: The extracted face profile and the absolute values of curvature.

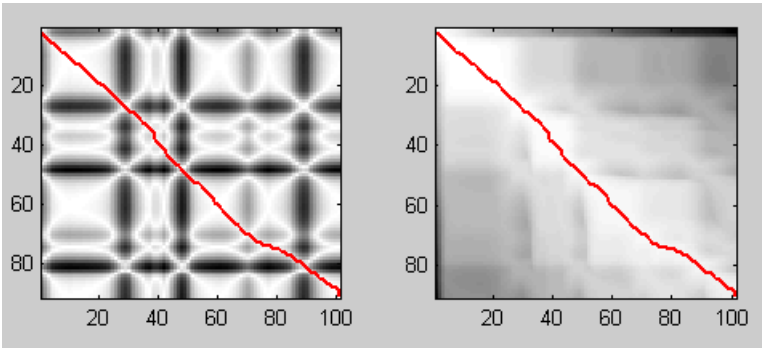

Figure 5: The similarity matrix (left) and the dynamic programming matrix (right).

of the aligned low-resolution images to obtain a more accurate high-resolution image. The Irani-Peleg method uses a point-spread function to model the optical and electronic properties of the image capture device used. We tested several filters to model the spatially variant point spread function (PSF) of the acquisition system in order to improve accuracy of the model. In this paper, we use a circular averaging filter with diameter 4.

The resolution of the original low-resolution face profile images is 70x70 and the resolution of the reconstructed high-resolution face profile image is 140x140. Figure 7 shows some of the low-resolution face profile images from adjacent video frames and Figure 8 shows the edges of these face profile images. Figure 9 shows the reconstructed high-resolution face profile image and its edges. For comparison, we resize the low-resolution face profile images by using bilinear interpolation. We can see that the reconstructed high-resolution image is much better than low-resolution images. It can be seen that in the edge images that the edges of the high-resolution image are much smoother and more reliable than that of the low-resolution images. In fact, for low-resolution images, we could not even obtain a complete face profile from the edge image. This explains why we need to apply super resolution algorithm to our problem. Using the reconstructed high-resolution image, we can extract more reliable features from face profile to perform the recognition task.

Recognition performance is used to evaluate the significance of our method, the quality of extracted features and their impact on identification. We distinguish two sequences for each person as sequence A and sequence B. From each sequence, we construct three high resolution face profile images using 33 frames, each of which uses 11 frames. We totally obtain 6 high-resolution images from both Sequence A and sequence B. Figure 10 shows these six high-resolution images for one person. We test recognition performance 6 times using different training images and testing images. For one half, we use the three high-resolution face profile images of each person from sequence A as training data, and each of the high-resolution face profile images from sequence B as the testing data in turn. For the other half, we use the three high-resolution face profile images of each person from sequence B as training data, and each of the high-resolution face profile images from sequence A as the testing data in turn. There are always 14 people and so 52 images in the training set and 14 images for 14 people in the testing set. We use 3-nearest neighbor to decide the label for a testing image. The label is the hypothesis, which gets the most vote based-on the first three similarity distance ranking from low to high. If the number of votes is equal, the label with the smallest distance is assigned to the testing image. A test image, which has a corresponding image in the database, is said to be correctly identified if the correct label is assigned. Table 1 shows the recognition result using our method. For the six experiments, the recognition result are different. It is so even for the different images constructed from the same sequence. It can be explained by the dynamics of a video, such as while people are moving and their face expressions may change. The best recognition rate is 78.6% (3 errors out of 14 per-

(a) (b) (c) (d)

Figure 6: Four adjacent video frames

(a) (b) (c) (d)

Figure 7: The four low-resolution side face images resized by using bilinear interpolation.

(a) (b) (c) (d)

Figure 8: The edge images of four low-resolution side faces

Figure 9: The reconstructed high-resolution side face and its edge image.

Table 1: Experimental results using our method

Testing Set	Training Set	
	Three images from sequence A	Three images from sequence B
Image 1 from sequence A		78.6%
Image 2 from sequence A		78.6%
Image 3 from sequence A		64.3%
Image 1 from sequence B	78.6%	
Image 2 from sequence B	71.4%	
Image 3 from sequence B	78.6%	

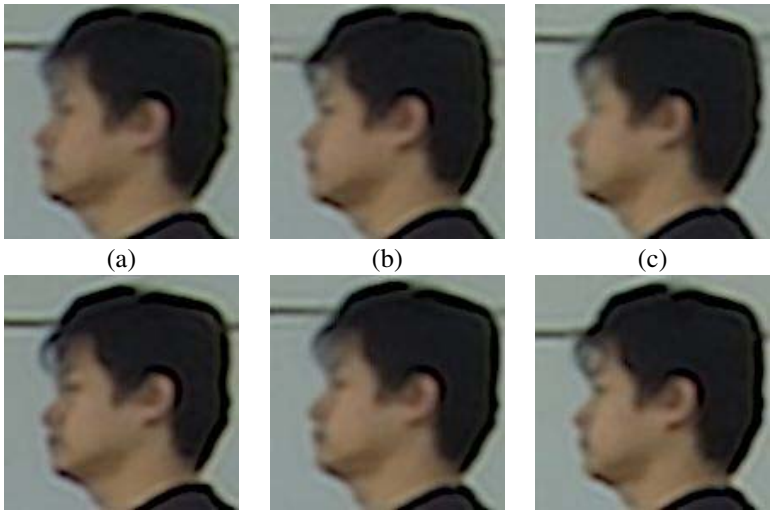

Figure 10: The six high-resolution side face images for one person. Each of the images is constructed from 11 frames.

sons) and the worst is 64.3% (5 errors out of 14 persons). On the average, more than 70% people are correctly recognized by face profile.

The experiments show that our approach for face profile recognition in video is promising. Since face profiles are sensitive to noise, when people are walking at a distance from a video camera, the face profiles shown in low-resolution videos are not reliable enough for the recognition task. The constructed high-resolution images have better quality and more reliable features from a face profile obtained. So our method is relatively robust in reality under different conditions. Although the experiments are done on a small database, our system has potential since it deals with real problems in a reasonable way and shows that face profile is a useful biometrics.

4 Conclusions

This paper introduces a practical approach for human recognition based on face profiles from a low-resolution video sequence. For optimal face profile recognition, we use both the spatial and temporal information presented in a number of consecutive frames to construct high-resolution face profile images. Some important issues that concern real-world applications are addressed in this paper. Experiments on a number of dynamic video sequences demonstrate the applicability and reliability of our method.

References

[1] J. Rosenfeld and Y. Regev, "Space-Time super resolution in video," VISL, Tech. Rep., 2004.

[2] S. Periaswamy and H. Farid, "Elastic registration in the presence of intensity variations," *IEEE Transactions on Medical Imaging*, vol. 22, no. 7, pp. 865–874, 2003.

[3] R. Tsai and T. Huang, "Multiframe image resoration and registration," in *Advances in Computer Vision and Image Processing*. JAI Press Inc., 1984.

[4] S. Borman and R. Stevenson, "Spatial resolution enhancement of low-resolution image sequences - a comprehensive review with directions for future research," University of Notre Dame, Tech. Rep., 1998.

[5] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Processing Magazine*, vol. 20, pp. 21–36, 2003.

[6] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion and transparency," *Journal of Visual Communication and Image Representation*, vol. 4, pp. 324–335, 1993.

[7] ——, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, pp. 231–239, 1991.

[8] B. Bhanu and X. Zhou, "Face recognition from face profile using dynamic time warping," in *17th International Conference on Pattern Recogntion*, vol. 4, 2004, pp. 499–502.