Facial Model Estimation (FME) Algorithms Using Stereo/Mono Image Sequence

Tsang-Gang Lin[a] and Chung J. Kuo[b]

[a]Opto-Electronics & Systems Laboratories, Industrial Technology Research Institute, Chutung, Taiwan 31040

[b]Institute of Communication Engineering, National Chung Cheng University, Chiayi, Taiwan 62107

ABSTRACT

Facial model generation is an important issue in the model-based applications, such as MPEG-4 and the virtual reality. An effective and precise construction algorithm of 3D facial model from 2D images should be available for practical applications. To generate facial model usually requires stereoscopic view of the face in the pre-processing stage. Although facial model can be successfully estimated from two stereo facial images, the occlusion effect and imprecise location of the feature point prohibit us from obtaining an accurate facial model. In this paper, we proposed several facial model estimation (FME) algorithms to find the precise facial model form a stereo or mono image sequence. The information of head movement, which is recorded in the image sequence, in the temporal domain is utilized for the facial model estimation. Even though the *a prior* information about the 3D position of the head with respect to the camera and the rotation axis and angle of the head's movement are unknown, an accurate facial model (within 7.21% error) can still be obtained by our schemes. In addition, our schemes do not require the precise camera parameters and avoid the tedious camera calibration such that the facial model generation is easily achieved.

Keywords: Facial model, Model generation, Model-based coding, Facial image sequence

1. INTRODUCTION

Very low bit rate coding of (stereo) video signals is very demanding due to the advent of visual communications and limit of communication channel.[1] For the applications of very low bit rate video coding, the (face and body) model-based coding defined in MPEG-4 standard[2] is a very promising scheme. It is known that facial image is the most important type of images within the video signals. Therefore, most model-based coding schemes are thus concentrated in facial images.[3]

In facial model coding, 3D model of a human face is first synthesized. 3D facial model can be extracted from stereo facial images from, at least, two perspective views which are usually obtained by using two cameras separated by a distance d_s.[4] The two cameras should be calibrated and their parameters should be obtained in order to calculate the facial model. Since camera calibration and camera parameters extraction are tedious works, the 3D facial model is thus difficult to obtain.[5] In addition, some occlusion effects also exist within a pair of stereo image. In this case, the position of the occluded feature points cannot be estimated precisely.

Several algorithms for adapting facial model have been proposed. A facial model is adapted manually by Aizawa and Saito.[6] Some recent research[7] show that a facial model can be obtained from frontal and side view images, and still some[8] show that the facial model can be derived from stereo image pair. In estimation of the 3D position from the corresponding points in the left- and right-view image frames, the mismatch due to the mis-correspondence creates large errors in the facial model. To solve these problems, we propose a facial model updating algorithm by using the redundancies which come form the motion and disparity of the stereo image sequence.

Practically most image sequences consist of mono-scopic but not stereo view. Here, we proposed other two algorithms to extract the facial model from a mono image sequence. The first one extract the model parameters from two consecutive image frames (at two different time), while the second one updates the facial model during the course of time in mono image sequence.

Further author information: (Send correspondence to Chung J. Kuo)
Chung J. Kuo: E-mail: kuo@ee.ccu.edu.tw

In *Input/Output and Imaging Technologies II*, Yung-Sheng Liu, Thomas S. Huang, Editors,
Proceedings of SPIE Vol. 4080 (2000) • 0277-786X/00/$15.00

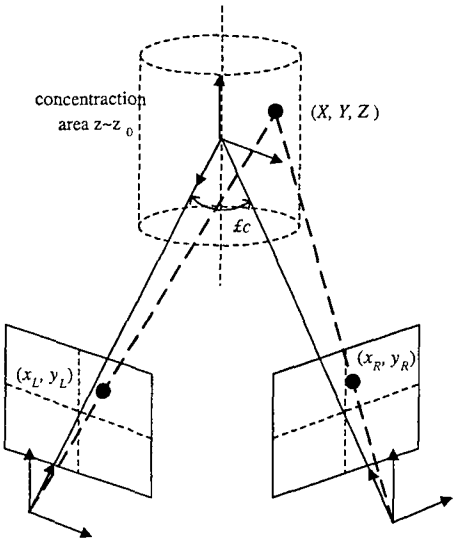

Figure 1. The geometry of the stereo imaging system

2. 3D HUMAN FACE MODELING

2.1. 3D Facial Modeling through Stereo Images

The estimation of 3D position in a stereo imaging system is accurate given the well-calibrated parameters, but the calibration of stereo imaging system is quite complicated. Chung and Nagata[9] proposed a camera model in which only the focal length f, the convergence angle θ (the angle between the optical axes of two cameras), and the baseline distance D (the distance between the centers of two cameras) are necessary. These parameters can be obtained by direct measurement or found in the camera data sheet.

As shown in the figure 1, the distances from both camera centers to the world coordinate origin are the same, and the value if z_0 can be simply driven using its internal parameters as

$$z_0 = \frac{D/2}{sin(\theta/2)} \tag{1}$$

If a point in the concentration area ($z \sim z_0$) is projected into the left and right image planes and their image plane coordinates are respectively (x_L, y_L) and (x_R, y_R), the coordinates of the point in the left camera frame are approximately determined as

$$X = \frac{z_0}{f} x_L \tag{2}$$

$$Y = \frac{z_0}{f} y_L \tag{3}$$

$$Z = \frac{z_0}{f} \frac{x_L cos\theta - x_R}{sin\theta} \tag{4}$$

2.2. 3D Facial Modeling through Single Front-view Image

The essential concepts of lateral parameter estimation can be described as follows. If we want to estimate the depth of mouth, θ, for one unknown person, we must first obtain the observation, Y, such as the width of mouth and several related parameters from his frontal face. The raw value of mouth width is converted into standardized score, Z_Y. Suppose this score is $1.1SD + mean$ in the population of our database, what will be the possible z-score of the lateral mouth depth? To solve such problem, we must obtain the conditional probability of Z_θ given $Z_Y = 1.1SD$,

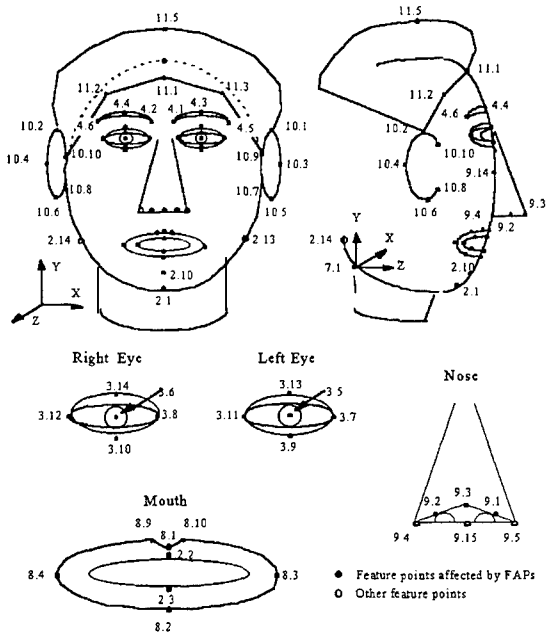

Figure 2. The feature points used in this study.

and estimate Z_θ by the previous discussed rules. We may obtain a set of estimates for Z_θ. Therefore, a criterion for selecting one as the best estimate must be made. Finally, we convert the estimated Z_θ into real value in metric, and thus we obtain the depth of mouth.

3. FACIAL MODEL ESTIMATION THROUGH STEREO IMAGE SEQUENCE

Feature extraction is the preliminary procedure for the advanced processing. We adopt the scheme including the template matching and correlation techniques similar to Nguyen and Huang.[10] The feature point used in this study is a sub-set shown in Fig. 2. Once all the feature points in the face are located from a stereo image pair, the facial model can be estimated by calculating the disparity between the feature points in the left and right image frame. Theoretically, in stereo image sequence, stereo matching is only necessary for the initial pair of frames. For the subsequent pairs, stereo correspondence can be found from the temporal motion vectors and stereo matching would be necessary only for the features that newly enter the field of view. However, because both temporal motion estimation and disparity estimation are ill-defined problem, few prior researches have been devoted to the fusion of stereo disparity and 3D motion estimation.[11,12] In reality, the 3D facial model obtained from the first pair of image frames in a stereo image sequence is not accurate. The reasons are due to the possible occlusion and the inaccuracy in the feature point's position, disparity and camera model. To solve this problems, we must include more information, if any, from the other pair of stereo images.

3.1. Off-Line Facial Model Estimation for Two Stereo Image Pairs

A facial model extraction (FME) algorithm for two pairs of stereo images is first proposed to solve these problems. Here we impose no constraint on the time distance between these two pairs of the stereo images and assume that the camera model is unknown. Since the time difference between the two pairs of stereo images may be large, the occlusion effects can be small and the model can be estimated more accurately. This proposed algorithm is not intended for real-time applications and thus named as off-line facial model estimation (off-line FME) algorithm.

Off-Line FME Algorithm

Assuming two pairs of stereo facial images are given. The images are $I_{t,s}$, where $t = 1, 2$ denotes the first and second pair of stereo images and $s = L, R$ denote the right or left view image frame. Thus we have four image frames, $I_{1,L}$, $I_{1,R}$, $I_{2,L}$, $I_{2,R}$. The relationship of disparity and 3D motion among these four image frames is illustrated in Fig. 3. In addition, we assume that the focal length of the lens in the imaging system and the distance and convergence angle between the two cameras are known. These are the parameters necessarily in a stereo imaging system. The following algorithm will estimate the facial model from these image frames if the perspective projection is used to model the image formation.

1. Locate the face and $FP_{1,L}$ (the position of feature point F) from image $I_{1,L}$. Here F can be any feature point shown in Figure 1. Use the feature point position $FP_{1,L}$ as an initial guess to find $FP_{1,R}$, $FP_{2,L}$ and $FP_{2,R}$ (the position of feature point F in image plane of $I_{1,R}$, $I_{2,L}$ and $I_{2,R}$ respectively).
2. Estimate the facial model (M_1, the 3D positions of all the feature points) according to the feature points' position ($FP_{1,L}$ and $FP_{1,R}$), the disparity and focal length of the imaging system.
3. Find the rotation angle and translation distance of model M_1 such that the 2D feature points' position in M_1 and image $I_{2,R}$ are best matched according to the perspective projection scheme.
4. Calculate the 2D perspective projective positions of all the feature points in M_1 in the right image plane after rotation and translation and denote them as $FP'_{2,R}$.
5. Averaging $FP'_{2,R}$ with $FP_{2,R}$ to obtain the new estimated feature points $\overline{FP}_{2,R}$.
6. Estimate the facial model (M_2) according to the feature points' position $FP_{2,L}$, estimated feature points ($\overline{FP}_{2,R}$), disparity, and focal length of the imaging system.
7. Repeat the above procedures for image frames $I_{1,L}$ and $I_{1,R}$ to obtain facial model M_3.
8. Similarly, we have the facial model sequence M_i, $i = 1, 2, \cdots$. The computation of the facial model stops when $M_i \approx M_{i+1}$.

3.1.1. Transformation of Facial Model

When the corresponding feature points of the stereo image pair, $I_{t,L}$ and $I_{t,R}$, are detected, their 3D positions can be reconstructed. The rough model M_t can thus be established. So does another model M_{t+1} which are obtained from $I_{t+1,L}$ and $I_{t+1,R}$. If the corresponding feature points are correctly extracted, M_t and M_{t+1} will be identical in this ideal case. Unfortunately, it is seldom to be so. The close-loop relationship of motion and disparity facilitates us to correct the errors derived from the mis-matched feature points. We will elaborate the this key technique of our model estimation algorithm in this section.

Analytically, the 3D motion of a rigid object can be decomposed as rotation and translation. Thus relation of M_t and M_{t+1} can be written as

$$M_{t+1} = \mathbf{R} \times M_t + \mathbf{T}, \tag{5}$$

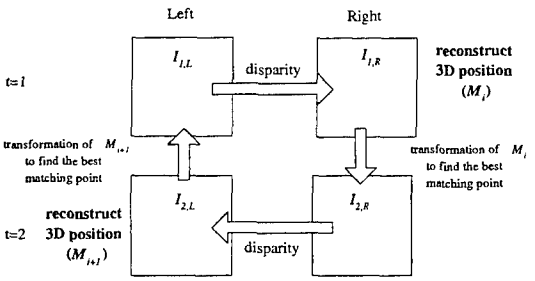

Figure 3. Illustration of off-line FME algorithm through stereo image pairs

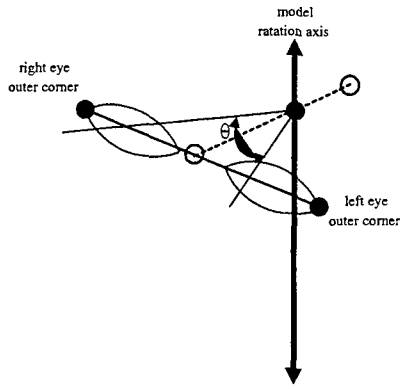

Figure 4. Transformation origin axis and rotation angle

where $\mathbf{R}$ is the rotation matrix and $\mathbf{T}$ is the translation matrix. Our purpose is to estimate the transformation , $\mathbf{R}$ and $\mathbf{T}$, of the head in the time interval. Firstly, M_t should be translated and rotated, so that the perspective projections , denoted as $FP'_{t+1,R}$, is fitting the corresponding feature points in $I_{t+1,R}$, denoted as $FP_{t+1,R}$. The translation is regarded as the displacement between the centroid of $FP'_{t+1,R}$ and the centroid of $FP_{t+1,R}$. It should be explained that if the translation in the Z coordinate exists, the head shown in the image frames of different time will have a scaling factor. But in our database, we impose no such condition. So the translation in the Z coordinate, T_Z, is set to be zero, and only the translations in the X and Y coordinate, T_X and T_Y, are calculated.

Choosing the correct location of rotation axis is required during rigid object transformation, but it is hard to make the decision without any priori knowledge about human motion mechanism. Thus a full search method is adopted to overcome the uncertainty. Here we set the rotation axis align with the middle point of the outer corner points of both eyes (feature points 3.7 and 3.12) in the vertical coordinate. But its depth is along the normal direction of the eye outer corners section toward the other side of the nose. The symmetry of human head is taken into consideration and the fact that the rotation axis locates in the inner skull is guaranteed. Figure 4 illustrate the relation of the eye corners and the rotation axis.

The variables in the full search method are the position of rotation axis and the rotation angle. The distance between the outer corners of both eyes is denoted as D_{corner}. Then the possible rotation axis position is located at the plane which is perpendicular to the outer eye corners connection and passes through the middle point, P_{mid}, of these two corners. Further, we consider the distance between the axis and P_{mid} is proportional to D_{corner}. We assume the range of the scaling factor is from 0 to 1, and the rotation angle is from $-2/\pi$ to $+2/\pi$. The resolution is adjusted hierarchically to achieve a higher accuracy and better extremum. The initial search resolution is 0.1 D_{corner} and 1°, and then increase the resolution to 0.01 D_{corner} and 0.1° and so on. Finally, the distance between the projections of M_t ($FP'_{t+1,R}$) and their corresponding points ($FP_{t+1,R}$) is used as a decision measure.

3.1.2. Convergence of Iteration Process

The relative position error is used to check the convergence of model during the iteration process. Respectively, let $F_{i,j}$ and C_i be the 3D position of the jth feature point and the centroid of all the feature point in model M_i at the ith iteration. The condition of convergence for our FME algorithm is

$$\frac{\sum_j [||F_{i,j} - C_i| - |F_{i+1,j} - C_i||]}{\sum_j |F_{i+1,j} - C_j|} \leq V_{th}, \tag{6}$$

where V_{th} is a pre-decided threshold value. The position error is the distance between the two corresponding feature points in the right and left view images. Due to the possible translation and rotation between the face in the first and second pair of stereo images, direct calculation of the position error suffers from the inherited bias term. To solve this problem, we thus make the centroid of all the feature points in the two models (M_i and M_{i+1}) coincides and use relative error as single measure.

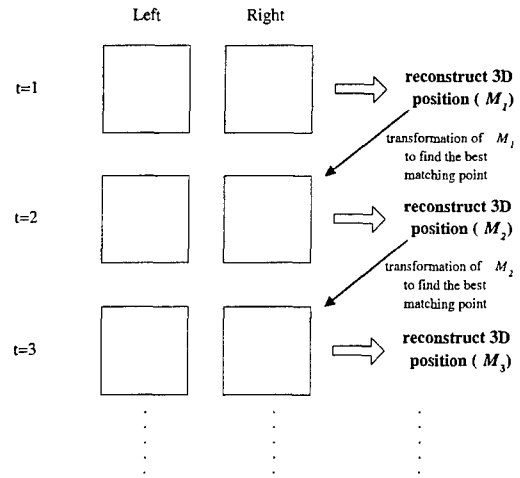

Figure 5. Illustration of on-line FME algorithm through a stereo image sequence

In summary, the proposed algorithm finds the 'global motion' of the head movement and the model parameters at the same time. Since the amount of global motion found (by any mean) will affect the precision of the model parameters, an algorithm is proposed to solve this problem iteratively. For two pairs of stereo images, the global motion between these two pairs is fixed and so is the model parameters. Therefore, the proposed algorithm will converge and the global motion and model parameters can be accurately estimated.

Throughout the development of this algorithm, the camera parameters we know are the focal length of the lens in the imaging system and the distance and convergence angle between the two cameras. The distance and convergence angle can be directly measured and the focal length can be obtained from the camera's data sheet. Therefore no camera calibration is required but we can only obtain the 'relative' facial model where the unit of feature point's position is pixel. If the relationship between the 'unit' in image plane and real world is known, then the exact facial model is obtained. To achieve this, camera calibration is necessary.

3.2. On-Line Facial Model Estimation for Stereo Image Sequence

Although the algorithm shown in the previous section is intended for off-line applications, it can be easily modified for on-line applications. That is, for stereo image sequence $I_{t,i}$, the off-line algorithm can be applied to images $I_{t,i}$, $t = 1, 2$ to obtain the estimated model. Subsequently, we have the second estimated model when $t = 2, 3$ and so on. However, to do so requires a high speed computer because the algorithm requires several iterations to find the model. To solve this problem, we modify the off-line algorithm such that it can be used for on-line applications.

On-Line FME Algorithm

1. Same as the step $1 \sim 7$ shown in the off-line FME algorithm for two stereo images.
2. Repeat the procedures for the image frames $I_{3,L}$ and $I_{3,R}$ to obtain facial model M_3.
3. Consecutively the following image pairs come in the processing. As a result, we have the facial model sequence M_i, $i = 1, 2, \cdots$ in the forward direction. The computation of the facial model stops when $M_i \approx M_{i+1}$.

Above algorithm is illustrated in Fig. 5.

According to the $FP_{1,L}$ and $FP_{1,R}$, the 3D position of facial model M_1 is estimated. In the previous section, Equation (5) demonstrates that a deformation relation exists between M_1 and M_2. We use a iterative full search method to explore the rotation angle and translation distance. This process is described in detail in the Subsection 3.1.3. Then the new estimated feature points $\overline{FP}_{2,R}$ is calculated by averaging the projective points of M_1 ($FP'_{2,R}$) in the right image plane and the original feature points ($FP_{2,R}$). A newly facial model M_2 is thus estimated according

to $FP_{2,L}$ and $\overline{FP}_{2,R}$. Repeating the procedure for M_2 and the next image pair $I_{3,L}$ and $I_{3,R}$, then estimating the facial model M_3. As the sequence goes on, we have the facial model sequence M_i, $i = 1, 2, \cdots$. If the convergence is achieved, which means that $M_i \approx M_{i+1}$, the computation of the facial model stops. The criterion of convergence, the same as one one in the off-line FME algorithm, is stated in Equation (6). Finally, a more accurate facial model $\overline{M}$ is concluded by averaging the facial model sequence. But each model is obtained at different temporal point, there are slight difference between adjacent models. Directive averaging is not correct. We have to adjust these models so that they have the same direction. Since the rotation angle and translation distance are known in the above procedure, a inverse computation is applied to enforce every models face the same direction. Then the final estimated facial model $\overline{M}$ is calculated as

$$\overline{M} = \frac{1}{N} \sum_i M_i. \tag{7}$$

4. FACIAL MODEL ESTIMATION THROUGH MONO IMAGE SEQUENCE

In last section, we discuss the two proposed algorithms that can estimate a more accurate facial model using off-line and on-line techniques. But these two FME algorithms require stereo image sequence, the utilizations are limited due to the inconvenience of stereo image capture. In an ordinary case, only a single-view image sequence is obtained using the normal image capture systems, such as CCD camera or traditional VCR. This makes us to design a new facial model estimation algorithm, in which the only input is the mono image sequence. In a stereo image sequence, the redundancy comes from both disparity and motion, but in a mono image sequence, only the motion information is shown in the image content. Thereby a challenge is implied and deserves our efforts. Inheriting form the concepts in the FME algorithm through stereo image sequence, we tend to design the FME algorithm through mono image sequence in two aspects, off-line and on-line. The following sections give the details about the algorithms and their implementation procedures .

4.1. Off-Line Facial Model Estimation for Two Mono Images

Here we impose no constraint on the time interval between these two images. Since the time interval between the two images may be large, the two images may have large discrepancy and thus the model can be estimated more accurately. This proposed algorithm is not intended for real-time applications and thus named as off-line facial model estimation (off-line FME) algorithm.

Assuming two facial images are given. The images are I_t where $t = 1, 2$ denotes the first and second image. The following algorithm will estimate the facial model from these two image frames. Here orthogonal projection is used to model image formation because we assume the camera is directly in front of the face and the first image (I_1) must contain the front view facial image. This implies the image plane (u,v) coincides with the $X - Y$ plane of the model coordinate. The depth information in the Z axis is what we want to estimate here.

Off-Line FME Algorithm

1. Locate the face and FP_1 (the position of feature point F) from image I_1. Here F can be any feature point shown in Figure 1. Use the feature point position FP_1 as an initial guess to find FP_2 (the position of feature point F in image plane of I_2).
2. Estimate the facial model (M_1, the 3D positions of all the feature points) according to the feature point's position (FP_1) and the anthropometric estimation scheme shown in.[13]
3. Find the rotation angle θ and location of rotation axis $l_{x,z}$ of model M_1 (with respect to its center) such that the 2D feature point's position from M_1 (after rotation) and image I_2 are best matched.
4. Rotate model M_1 by θ and then calculate the 2D positions of all the feature points (through projection) that are denoted as FP_2'. $\overline{FP}_2$ is defined as the average of FP_2 and FP_2'.
5. Combine the $\overline{FP}_2$ and the depth of each feature point of the rotated model M_1 to construct M_{temp}, the 3D position at second image time instance.
6. Repeat Step 3 such that the 2D feature point's position from (M_{temp}) (after projection) and FP_1 are best matched. Then use the idea in Steps 4 and 5 to find M_2.

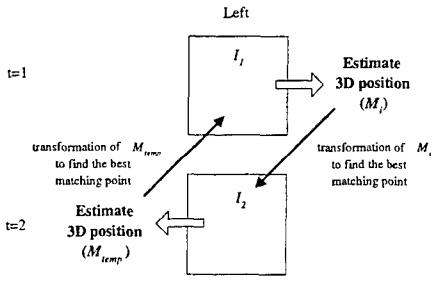

Figure 6. Illustration of off-line FME algorithm through a mono image sequence

7. Keep looping the above procedures for images at $t = 1$ and $t = 2$, we have facial model sequence M_i, $i = 1, 2, \cdots$. The computation of the facial model stops when $M_i \approx M_{i+1}$.

Figure 6 shows the idea of this algorithm.

3D motion of a rigid body consists of rotation and translation. The relationship between the facial model M_i and M_{temp} is

$$M_{temp} = \mathbf{R} \times M_i + \mathbf{T}, \tag{8}$$

where $\mathbf{R}$ and $\mathbf{T}$ are the rotation and translation matrix, respectively. According to the relationship between model M_i and M_{temp}, we have

$$x_{temp} = x_i \cos\theta + z_i \sin\theta + T_x \tag{9}$$

$$y_{temp} = y_i \tag{10}$$

$$z_{temp} = -x_i \sin\theta + z_i \cos\theta + T_z \tag{11}$$

where (x_i, y_i, z_i) is the 3D position of the feature point F. Because the rotation axis is parallel to the Y axis, y_i and y_{temp} should keep the same. Since the orthogonal projection is employed in monoscopic case, we have the 2D position of the feature point F in I_1 and I_2 as (x_1, y_1) and $(x_1 \cos\theta + z_1 \sin\theta, y_1)$, respectively, where no head translation exists between these two time instances.

Since I_1 contains the front view facial image, we can use the scheme shown in[13] to estimate the depth information z_1. Once the z_1, location of rotation axis $l_{x,z}$, rotation angle θ, and the 2D position of feature point F at time $t = 1, 2$ are known, we can update the position of feature point (in x and z coordinate) according to Equations ?–?. By the iteration process, an almost accurate position of the feature point F can thus be obtained.

4.2. On-Line Facial Model Estimation for Mono Image Sequence

If the input image sequence is plenty enough, the on-line FME algorithm is easily obtained after applying slight modifications from the off-line FME algorithm. The only difference between these two algorithms is at the step when the facial model of the second image is estimated. Off-line algorithm returns backward to the first image, but on-line algorithm seeks forward to the next image. The same problem comes from the different facing direction in the sequence content is addressed out and has to be solved.

On-Line FME Algorithm

Although the algorithm shown above is intended for off-line applications, it can be easily modified for on-line applications. That is, it is first applied to images I_1 and I_2. Then images I_2 and I_3, and so on. However, to do so needs a high speed computer because of the amount of computations required. To solve this problem, we modify the original algorithm for on-line processing. Considering the mono image sequence I_t where $t = 1, 2, \cdots$, the on-line FME algorithm is shown below.

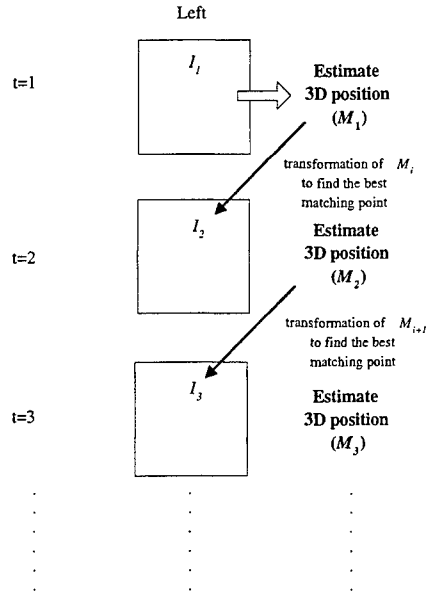

Figure 7. Illustration of on-line FME algorithm through a mono image sequence

1. Same as the step $1 \sim 4$ shown in the off-line FME algorithm for two mono images.
2. Combine the $\overline{FP}_2$ and the depth of each feature point of the rotated model M_1 to construct M_2, the 3D position at second image time instance.
3. Repeat the above procedure to update model M_i from image I_i, $i = 3, 4, \cdots$. The computation of the facial model stops when $M_i \approx M_{i+1}$.

Figure 7 shows the idea of this algorithm.

5. RESULTS

In all the four algorithms proposed in this paper, we can only obtain the 'relative' facial model where the unit of feature point's position is 'pixel' but not '*mm*.' Therefore, a scaling factor must be found beforehand for accuracy evaluation. In this section the estimated value from our FME algorithm is attached with a hat, while the actual with no accessory, such as $\hat{F}_j$ and F_j respectively.

Since the rotation axis, parallel to Y axis, was found in model estimation, we first define the shortest distance between the jth feature point (F_j) and the rotation axis $l_{x,z}$ as $d_{xz}(F_j)$. Then the scaling factor (SF) of the feature point with respect to the rotation axis of the head is

$$SF_{xz} \equiv \frac{1}{N} \sum_j \frac{d_{xz}(F_j)}{d_{xz}(\hat{F}_j)} \tag{12}$$

where the unit of this scaling factor is mm/pixel.

Similarly, we define the vertical distance between a feature point (F) and the topmost feature point 3.13 (or 3.14) in the facial model as $d_y(F)$. The scaling factor of the feature point with respect to the feature point 3.13 (or 3.14) is

$$SF_y \equiv \frac{1}{N} \sum_j \frac{d_y(F_j)}{d_y(\hat{F}_j)} \tag{13}$$

The overall scaling factor between the actual and estimated facial model is thus defined as the average of SF_{xz} and SF_y. Therefore we can scale the estimated model such that it is about the same size as the actual model. Then the estimated model after scaling is denoted as $\hat{F}_j{}'$, and its unit is '*mm*.'

To know the accuracy of the model estimated, single measure must be defined. Respectively, let F_j and C be the actual 3D position of the jth feature point and rotation center and $\hat{F}_j{}'$ and $\hat{C}'$ denote the estimated 3D position (in *mm*). The relative error in the estimated facial model is then defined as

$$Error \equiv \frac{1}{N} \sum_j \frac{[||F_j - C| - |\hat{F}_j{}' - \hat{C}'||]}{|F_j - C|}. \tag{14}$$

Table 1 shows the facial model obtained from both off-line and on-line FME algorithms for stereo image sequence. The results convince us that the correct calibration of camera parameters is not important and an added advantage for the use of two pairs stereo images is to lessen the necessity of camera calibration. In addition, the redundant information in the image sequence helps to generate a improved facial model with less error.

The facial model derived from the mono image sequence is shown in Table 2, where the relative positions of all the feature points are listed. Comparing with the results shown before, the on-line algorithm still finds the accurate position of all the feature points (within 6.56% error). In addition, for mono images, on-line algorithm provides results with smaller error compared with the off-line algorithm because of the additional information available in the image sequence.

According to the accurate positions of all necessary feature points, a 3D mesh model is thus generated. Then the first image I_1, which is the front-view facial image under the beginning assumption, is applied as a texture material in the texture mapping procedure. The mapping method is the UV plane mapping. Figure 8 shows the estimated facial model (after texture mapping) from two mono image images. Although this is the largest error feature point set, the results resemble true face very well. And the other three algorithms show the better performance.

6. CONCLUSION

Extraction of facial model from stereo image sequence is important for MPEG-4 related applications. Conventional approaches estimate the facial model from only one pair of stereo images but suffer from the occlusion effects and tedious camera calibration. In this paper, we propose the off- and on-line facial model estimation algorithms to estimate the facial model from stereo image sequence. Simulation results show that our schemes can obtain a much more accurate facial model compared with the conventional approach without the camera calibration.

Our schemes are suitable for practical on- and off-line applications. For off- line applications, two pairs of stereo facial images are first captured and our algorithm can be used to obtain the facial model without camera calibration. With slight modifications of the off-line FME algorithm, our algorithm can also be used for on-line facial model estimation.

Moreover,we extend the application to the mono image sequence. It was shown in[13] that facial model can be estimated from single front-view facial image. However, some model parameters (that is, the positions of some feature points) suffer large estimation error. Our subsequent works concentrate on the modification of the proposed algorithm here for mono image sequence. Inherented from the concepts of on- and off-line in the stereo case, both algorithms are established. The accuracy of the proposed FME algorithms through mono image sequence depends a lot on the initial facial model estimated from the first image. As a consequence the first image comes into the mono FME algorithms has to be chosen carefully. The lower error results convince us that the proposed facial model estimation algorithms through a mono image sequence are practicable.

REFERENCES

1. R. Schäfer & T. Sikora, "Digital video coding standards and their role in video communication," *Proceedings of the IEEE*, vol. 83, pp. 907–924, June 1995
2. "MPEG-4 Overview (Tokyo Version)," *Coding of moving pictures and audio*, ISO/IEC JTC1/SC29/ WG11 N2196, Mar. 1998

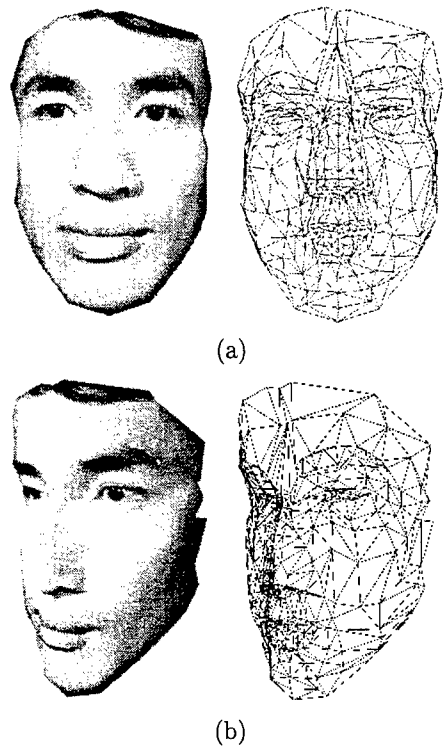

Figure 8. The estimated facial model from two mono images. (a)frontal view, and (b)45-degree view

3. S.C. Chang, K. Aizawa, H. Harashima & T. Takebe, "Analysis and synthesis of facial image sequences in model-based image coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 4, pp. 257-275, Jun. 1994
4. G. Galicia & A. Zakhor, "Depth based recovery of human facial features from video sequences," *Proceedings of IEEE International Conference on Image Processing*, pp. 603–606, Oct. 1995
5. R.C. Gonzalez & R.E. Woods, *Digital Image Processing*, Reading: Addison- Wesley, 1992
6. K. Aizawa, T. Saito & H. Harashima, "Model-based analysis synthesis image coding (MBSAIC) system for a person's face," *Journal of Signal Processing: Image Communication*, vol. 1, pp. 139–152, Oct. 1989
7. H. Tao, and T. S. Huang, "Deriving Facial Articulation Models from Image Sequences," *Proceedings of IEEE International Conference on Image Processing*, October 1998.
8. I. A. Kakadiaris and D. Metaxas, "Model-Based Estimation of 3D Human Motion with Occlusion Based on Active Multi-Viewpoint Selection," *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 81–87, June 1996
9. J.M. Chung & T. Nagata, "Binocular vision planning with anthropomorphic features for grasping parts by robots," *Robotica*, vol. 14, pp. 269– 279, 1996
10. T. Nguyen, T. Huang, "Segmentation, grouping and feature detection for face image analysis," *Proceedings of IEEE International Symposium on Computer Vision*, pp. 593–598, Nov. 1995
11. X. Chen & A. Luthra, "MPEG-2 multi-view profile and its application in 3DTV," *SPIE Proceedings*, vol. 3021, pp. 212–223, 1997
12. W. Richard, "Structrue form stereo and motion," *Journal of the Optical Society of America A*, vol. 2, pp. 343–349, Feb. 1985
13. C.J. Kuo & R.S. Huang, "Synthesizing lateral face from frontal facial image using human anthropometric estimation," *Proceedings of IEEE International Conference on Image Processing,* vol. 1, pp. 133–136, Oct. 1997

Table 1. Feature point's positions estimated from the proposed FME algorithm through stereo image suquence.

Feature Point	Off-Line FME Algorithm	On-Line FME Algorithm
3.12	(-94.073 -60.022 195.38)	(-96.87 -60.037 197.03)
3.14	(-62.504 -73.052 199.14)	(-63.97 -72.85 203.57)
3.8	(-35.005 -58.284 196.24)	(-36.368 -58.489 201.34)
3.10	(-63.651 -54.208 202.46)	(-63.957 -54.392 205.52)
3.6	(-63.075 -66.904 200.79)	(-63.942 -66.707 207.19)
3.11	(40.235 -62.093 196.31)	(40.939 -61.914 199.18)
3.13	(67.457 -74.789 194.22)	(68.802 -74.207 198.5)
3.7	(99.858 -59.153 192.48)	(99.498 -58.875 195.71)
3.9	(67.302 -57.015 197.75)	(69.555 -56.583 201.26)
3.5	(66.311 -67.907 197.55)	(68.569 -67.53 201.03)
8.4	(-51.181 78.165 200.12)	(-51.139 78.37 204.4)
8.9	(-20.651 59.522 234.6)	(-18.039 59.318 232.74)
8.1	(-6.1504 64.2 238.66)	(-2.4026 64.206 233.77)
8.10	(7.4639 59.322 239.15)	(11.021 59.298 233.41)
8.3	(54.356 79.168 210.72)	(55.481 79.122 211.34)
8.2	(-0.1476 108.77 227.39)	(2.3189 108.98 229.43)
2.2/2.3	(-4.4316 79.969 236.75)	(-1.2229 80.01 232.95)
9.4	(-42.211 27.648 207.93)	(-41.266 27.386 209.53)
9.3	(-16.832 16.756 269.69)	(-9.9246 16.032 255.48)
9.5	(41.061 23.104 221.62)	(42.896 22.203 222.25)
9.15	(-8.1349 36.803 244.43)	(-4.6879 36.657 238.8)
Error	4.6011%	3.9459%

Table 2. Feature point's positions estimated from the proposed FME algorithm through mono image suquence.

Feature Point	Off-Line FME Algorithm	On-Line FME Algorithm
3.12	(-93.92,-60.1,187.2)	(-95.6,-59.38,184.2)
3.14	(-63.63,-72.4,231.9)	(-66.14,-71.62,231.1)
3.8	(-34.62,-58.03,163.6)	(-36.76,-57.34,166.9)
3.10	(-62.63,-54.04,215.1)	(-64.92,-53.26,218.4)
3.6	(-61.13,-67.16,189.6)	(-63.15,-66.52,198)
3.11	(38.35,-62.02,164.4)	(35.48,-61.42,167.4)
3.13	(63.01,-74.26,191.1)	(59.93,-73.66,198.2)
3.7	(93.39,-57.97,187.4)	(90.03,-57.67,187)
3.9	(62.38,-56.81,229.2)	(58.78,-56.32,234.1)
3.5	(60.79,-67.16,240.5)	(57.06,-66.52,244.1)
8.4	(-50.27,77.77,196.3)	(-52.71,76.98,199.9)
8.9	(-10.71,59.11,234.9)	(-14.05,58.62,237.1)
8.1	(6.159,64.33,203.5)	(3.212,63.72,209.5)
8.10	(18.27,59.24,226.6)	(14.87,58.62,228.1)
8.3	(54.82,78.74,207.4)	(51.6,78,209.9)
8.2	(6.097,108.6,254.8)	(2.072,107.6,255)
2.2/2.3	(5.409,79.75,240)	(1.975,79.02,241.9)
9.4	(-40.82,27.22,234.8)	(-43.85,26.99,234.1)
9.3	(2.333,16.67,285.4)	(-1.855,16.45,282)
9.5	(45.08,21.95,225.4)	(41.49,21.55,227.9)
9.15	(4.903,36.52,229.2)	(1.624,36.17,232.2)
Error	7.21%	6.56%