A Probabilistic Network for Facial Feature Verification

Kyoung-Ho Choi, Jae-Joon Yoo, Tae-Hyun Hwang,
Jong-Hyun Park, and Jong-Hoon Lee

ABSTRACT— In this paper, we present a probabilistic approach to determining whether extracted facial features from a video sequence are appropriate for creating a 3D face model. In our approach, the distance between two feature points selected from the MPEG-4 facial object is defined as a random variable for each node of a probability network. To avoid generating an unnatural or non-realistic 3D face model, automatically extracted 2D facial features from a video sequence are fed into the proposed probabilistic network before a corresponding 3D face model is built. Simulation results show that the proposed probabilistic network can be used as a quality control agent to verify the correctness of extracted facial features.

I. INTRODUCTION

For various multimedia applications, such as video conferencing, e-commerce, and virtual anchors, talking heads are becoming more and more important to enriching the human-computer interface. To provide talking head solutions for these multimedia applications, which do not require high quality animation, researchers have investigated fast and easy ways to build a 3D face model to generate many different face models in a short time period. However, user intervention is still required to provide several corresponding points in two frames from a video sequence or feature points in a single frontal image [1]-[3].

To build a 3D or virtual face without any user intervention, it is highly desirable to have a robust decision algorithm check the correctness of extracted facial features before the 3D face model is built; this avoids the creation of unnatural or unrealistic 3D faces. This paper introduces a probabilistic approach for a robust decision algorithm for human frontal faces. In the proposed approach, we maximally use facial feature evidence to check the correctness of extracted facial features in a systematic way and build a 3D face model from a video sequence without any user intervention.

Manuscript received Sept. 30, 2002.

Kyoung-Ho Choi (phone: +82 42 860 5803, e-mail: khchoietri@etri.re.kr), Jae-Joon Yoo (e-mail: jjryu@etri.re.kr), Tae-Hyun Hwang (e-mail: hth63339@etri.re.kr), Jong-Hyun Park (e-mail: jhp@etri.re.kr), Jong-Hoon Lee (e-mail: jong@etri.re.kr) are with Spatial Information Technology Center, ETRI, Daejeon, Korea.

II. PROBABILITY NETWORKS FOR FACIAL FEATURE EXTRACTION

The proposed probabilistic network is a simple case of directed graphical models, also called Bayesian Networks [4], which can be used to describe the conditional independence relationship between variables, where deterministic relationships between nodes can be modeled using an arc. Probabilistic approaches can be applied to solving emerging AI problems, e.g., iris recognition and video summarization [5], [6], in a systematic fashion and have been successfully used to locate human faces from a scene and to track deformations of local features [7], [8]. Cipolla et al. [8] proposed a probabilistic framework to combine different facial features and face groups and achieved a high confidence rate for face detection from a complicated scene. Huang et al. [7] used a probabilistic network for local feature tracking by modeling locations and velocities of selected feature points. For our automated system, we adopt a probabilistic framework, which is similar to the work done by [8], to maximally use facial feature evidence for determining the correctness of extracted facial features before a 3D face model is built. Our approach is different from [8] in that we chose the distance between two feature points as a random variable for each node to model a human frontal face. In a study of face anthropometry [9], data were collected by measuring distances and angles among selected key points, e.g., corners of eyes, mouth, and ears, from a face to describe the variability of the face. We characterize a frontal face by measuring the distances and covariance between key points chosen from that study. We propose the Facial Feature Net, Face Shape Net, and Topology Net to verify the correctness of extracted facial features; this also enables the algorithm to extract facial features more accurately.

1. Network Hierarchy

Figure 1 shows the network hierarchy used in our approach: a Facial Feature Net with subnets Mouth Net and Eye Net, a Face Shape Net, and a Topology Net. In the network each node represents a random variable and each arrow denotes a conditional dependency between two nodes.

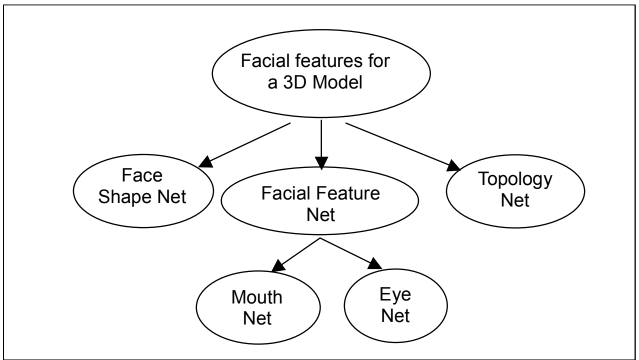

Fig. 1. The network hierarchy used in our approach.

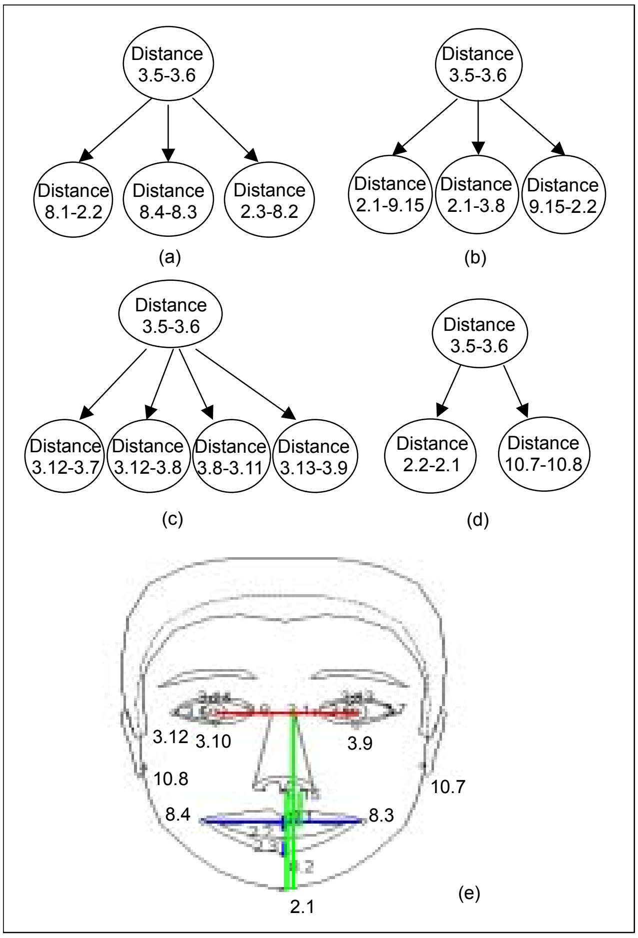

Fig. 2. The detail of each probability node. The distance 3.5-3.6 means the distance between FDP 3.5 and FDP 3.6 shown in (e): (a) Mouth Net, (b) Topology Net, (c) Eye Net, (d) Face Shape Net, and (e) graphical illustration of nodes in (a) and (b).

2. Modeling Probability Networks

All nodes in the proposed probability network are classified into four groups: Mouth = [D(8.1, 2.2), D(8.4, 8.3), D(2.3, 8.2)], Eyes = [D(3.12, 3.7), D(3.12, 3.8), D(3.8, 3.11), D(3.13, 3.9)], Topology = [D(2.1,9.15), D(2.1,3.8), D(9.15,2.2)], and Face Shape = [D(2.2,2.1), D(10.7,10.8)], where D(P1,P2) is a distance between facial definition parameters (FDPs) P1 and P2 defined in MPEG-4 standard [10]. In our network the distance between two feature points selected from the MPEG-4 facial object is defined as a random variable for each node. We model the distribution of random variables of two nodes.

For instance, we model D(3.5, 3.6), the distance between the centers of the left and right eyes, and D(2.1, 9.15), the distance of the two selected points FDP 2.1 and FDP 9.15, (Fig. 2(b)), as a 2D Gaussian distribution, estimating means, standard deviations, and correlation coefficients. Figure 2(e) shows graphical illustrations of the relationship between two nodes in the proposed probability network. For example, the length of the red line in Fig. 2(e), which is the distance between FDP 3.5 and FDP 3.6, and the length of the blue line, e.g., between FDP 8.4 and FDP 8.3 (width of mouth), are modeled as a 2D Gaussian distribution (1), where D_1, μ_{D_1}, and σ_{D_1} denote the distance between two selected FDPs, the means, and standard deviation of D_1, respectively, and ρ denotes the correlation coefficients between two nodes D_1 and D_2. To model the 2D Gaussian distributions of D(3.5, 3.6) and the distances of selected paired points, a database from [11] is used in our simulations. The reason we model probability distributions based on FDP 3.5 and FDP 3.6 is that according to our implementation, the left and right eye centers are the features

$$f_{pdf}(D_1, D_2) = \frac{1}{2\pi\sigma_{D_1}\sigma_{D_2}\sqrt{1-\rho^2}} \cdot \exp\left\{ -\frac{1}{2(1-\rho^2)} \cdot \left(\frac{(D_1-\mu_{D_1})^2}{\sigma_{D_1}^2} - \frac{2\rho(D_1-\mu_{D_1})(D_2-\mu_{D_2})}{\sigma_{D_1}\sigma_{D_2}} + \frac{(D_2-\mu_{D_2})^2}{\sigma_{D_2}^2} \right) \right\}, \tag{1}$$

that can be detected most reliably and accurately from a video sequence.

3. Calculating Probability

The chain rule and conditional independence relationship are applied to calculate the joint probability of each network. For instance, the probability of the Face Shape Net is defined as a joint probability of all three nodes, D(3.5, 3.6), D(2.2, 2.1), and D(10.7, 10.8):

$$\begin{aligned} P(\text{Face Shape Net}) &= P(\,D(3.5,3.6), D(2.2,2.1), D(10.7,10.8)\,) \\ &= P(\,D(3.5,3.6)\,)\times P(\,D(2.2,2.1)\mid D(3.5,3.6)\,) \\ &\quad \times P(\,D(10.7,10.8)\mid D(2.2,2.1), D(3.5,3.6)\,) \\ &= P(\,D(3.5,3.6)\,)\times P(\,D(2.2,2.1)\mid D(3.5,3.6)\,) \\ &\quad \times P(\,D(10.7,10.8)\mid D(3.5,3.6)\,). \end{aligned} \tag{2}$$

In the same manner, probabilities of other nodes can be defined as follows.

$$\begin{aligned} P(\text{Eye Net}) = {} & P(\,D(3.5,3.6)\,)\times P(\,D(3.12,3.7)\mid D(3.5,3.6)\,) \\ & \times P(\,D(3.12,3.8)\mid D(3.5,3.6)\,)\times P(\,D(3.8,3.11)\mid D(3.5,3.6)\,) \\ & \times P(\,D(3.11,3.9)\mid D(3.5,3.6)\,). \end{aligned} \tag{3}$$

$$\begin{aligned} P(\text{Mouth Net}) = {} & P(\,D(3.5,3.6)\,)\times P(\,D(8.4,8.3)\mid D(3.5,3.6)\,) \\ & \times P(\,D(8.1,2.2)\mid D(3.5,3.6)\,)\times P(\,D(8.2,2.3)\mid D(3.5,3.6)\,). \end{aligned} \tag{4}$$

$$P(\text{Facial Feature Net}) = P(\text{Eye Net})\times P(\text{Mouth Net}). \tag{5}$$

$$\begin{aligned} P(\text{Topology Net}) = {} & P(\,D(3.5,3.6)\,)\times P(\,D(2.1,9.15)\mid D(3.5,3.6)\,) \\ & \times P(\,D(2.1,3.8)\mid D(3.5,3.6)\,)\times P(\,D(9.15,2.2)\mid D(3.5,3.6)\,). \end{aligned} \tag{6}$$

$$\begin{aligned} & P(\text{Facial Feature, Face Shape, Topology}) \\ & = P(\text{Facial Feature Net})\times P(\text{Face Shape Net})\times P(\text{Topology Net}). \end{aligned} \tag{7}$$

In our implementation, P(Face Shape Net) is used to verify a face shape extracted from our face shape extractor, and P(Mouth Net) is used to check extracted mouth features. P(Topology Net) is used to determine if facial components, i.e., eyes, nose, and mouth, are located correctly along the vertical axis. P(Facial Features, Face Shape, Topology) of (7) is used as a decision criterion for the correctness of extracted facial features for building a 3D face model.

III. EXPREMENTAL RESULTS

We used 50 frontal face images of the PICS database from the University of Stirling (http://pics.psych.stir.ac.uk/) and the Expectation Maximization algorithm to model the proposed probabilistic network. We tested the proposed probability network as a quality control agent in our automatic system that creates a 3D face model from a video sequence without any

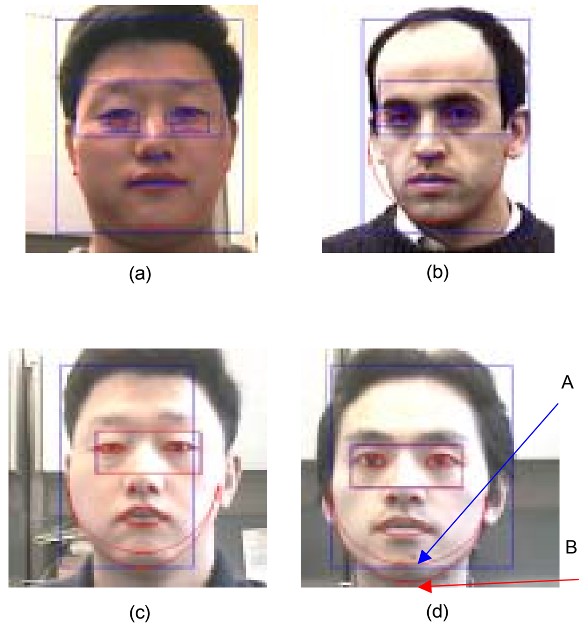

Fig. 3. The probabilistic network as a quality control agent: (a),(b) examples of rejected facial features; (c),(d) examples of increased extraction accuracy (contour A is an old face shape and contour B is one newly extracted by adjusting threshold values).

accuracy for facial feature extraction [12]. For our experiments, users were required to provide a frontal view with a rotation angle of less than ± 5 degrees. We recorded 20 video sequences, making a total of approximately 2000 frames. The face shape extractor achieved a detection rate of 64% for 1180 selected frames from the testing video sequences.

Figures 3(a) and (b) show examples of facial features rejected from the probabilistic network to prevent the creation of unrealistic faces. T_{fs}, the threshold value for face shape extraction, was adjusted automatically to improve accuracy based on the results of the probabilistic network ([12] contains the details of this procedure). If only P(Face Shape Net) was low, T_{fs} was increased to find a more clear boundary of the face. Figures 3(c) and (d) show examples of feature extraction improved by adjusting threshold values. According to the simulation results, the proposed probabilistic network was successfully combined with our automatic system to create a 3D face model, classifying 82.8% of the extracted facial features correctly. Slight errors in extracted eyes and mouth locations caused the errors in our experiments. Figure 4 shows examples of successfully created 3D face models. By using the probabilistic network approach, the chance of creating unrealistic faces due to wrong facial features was reduced significantly. For detailed experimental results, a demonstration of the talking head system is available at [13].

Fig. 4. Example of successfully created 3D face models.

IV. CONCLUSIONS

We have presented a probabilistic network to maximally use facial feature evidence in determining if extracted facial features are suitable for creating 3D face models. The distance of the two selected points from FDPs was chosen as a random variable for each node in our probabilistic network. We proposed the Facial Feature Net, Face Shape Net, and Topology Net to successfully verify the correctness of the extracted facial features, which also enable the algorithm to extract facial features more accurately. The proposed probabilistic network can be used as a quality control agent to verify the correctness of extracted facial features for various face detection and extraction systems.

REFERENCES

[1] Zicheng Liu, Zhengyou Zhang, Chuck Jacobs, and Michael Cohen, "Rapid Modeling of Animated Faces From Video," *Technical Report MSR-TR*-2000-11.

[2] C.J. Kuo, R.-S. Huang, and T.-G. Lin, "3-D Facial Model Estimation From Single Front-View Facial Image," *IEEE Trans. on CSVT*, vol. 12, no. 3, 2002, pp. 183-192.

[3] Ana C. Andres del Valle and Jorn Ostermann, "3D Talking Head Customization by Adapting a Generic Model to One Uncalibrated Picture," *IEEE Int'l Symp. on Circuits and Systems*, 2001, pp. 325-328.

[4] S. Russell and P. Norvig, *Introduction to Artificial Intelligence*, Prentice Hall, 1995.

[5] Jae-Gon Kim, Hyun Sung Chang, Young-tae Kim, Kyeongok Kang, Munchurl Kim, Jinwoong Kim, and Hyung-Myung Kim, "Multimodal Approach for Summarizing and Indexing News Video," *ETRI J.*, vol. 24, no. 1, Feb. 2002, pp. 1-11.

[6] Shinyoung Lim, Kwangyong Lee, Okhwan Byeon, and Taiyun Kim, "Efficient Iris Recognition through Improvement of Feature Vector and Classifier," *ETRI J.*, vol. 23, no. 2, June 2001, pp. 61-70.

[7] Hai Tao, Lopez R., and Thomas Huang, "Tracking Facial Features Using Probabilistic Network," *Automatic Face and Gesture Recognition*, 1998, pp. 166-170.

[8] K.C. Yow and R. Cipolla, "A Probabilistic Framework for Perceptual Grouping of Features for Human Face Detection," *Proc. of IEEE Int'l Conf. on Auto. Face and Gesture Recognition*, 1996, pp. 16-21.

[9] Lesilie G. Farkas, *Anthropometry of the Head and Face*, Raven Press, 1994.

[10] ISO/IEC FDIS 14496-2 Visual, ISO/IEC JTC1/SC29/WG11 N2502, Nov. 1998.

[11] *Psychological Image Collection at Stirling* (PICS), Available at http://pics.psych.stir.ac.uk/.

[12] K.H. Choi and Jenq-Neng Hwang, "A Real-Time System For Automatic Creation of 3D Face Models from a Video Sequence," *IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing*, 2002, pp. 2121-2124.

[13] http://allison.ee.washington.edu/~ khchoi/TalkingHeadSystem.html.