Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/092L21

Emotional Avatars: Appearance Augmentation and Animation based on Facial Expression Analysis

Taehoon Cho, Jin-Ho Choi, Hyeon-Joong Kim and Soo-Mi Choi∗

Department of Computer Science and Engineering, Sejong University, 98 Gunja, Gwangjin, Seoul 143-747, Korea

Received: 22 May 2014, Revised: 23 Jul. 2014, Accepted: 24 Jul. 2014
Published online: 1 Apr. 2015

Abstract: We propose an emotional facial avatar that portrays the user's facial expressions with an emotional emphasis, while achieving visual and behavioral realism. This is achieved by unifying automatic analysis of facial expressions and animation of realistic 3D faces with details such as facial hair and hairstyles. To augment facial appearance according to the user's emotions, we use emotional templates representing typical emotions in an artistic way, which can be easily combined with the skin texture of the 3D face at runtime. Hence, our interface gives the user vision-based control over facial animation of the emotional avatar, easily changing its moods.

Keywords: Emotional avatars, 3D facial animation, facial expression analysis, appearance augmentation

1 Introduction

The human face is a great communication medium because the face can evoke diverse facial expressions according to the internal emotions. One can read a person's emotional state from his/her facial expression and respond to it appropriately. In the cyberspace, a facial avatar that portrays specific emotions plays a significant role especially in non-verbal communication by improving its interpretation [1,2]. Such an emotional avatar has some advantage compared to a real face since it can offer another range of tone or feeling by augmenting the appearance of the virtual face with a variety of textured patterns. Hence, creating an emotional avatar requires an automatic augmenting approach where the avatar's appearance automatically changes according to a user's facial expressions while expressing his/her emotion creatively. However, most methods in computer graphics have focused on achieving realistic facial modeling and animation based on 3D face scans [3,4] or facial motion capture data [5,6].

To express emphasized emotions using a 3D facial avatar, we present a framework in which facial expression recognition is integrated with 3D facial animation. Our aim is to allow a user to freely make emotional facial expressions in front of a webcam, and analyze the facial expressions to animate a 3D facial avatar with an artistic style emphasizing emotions. The overview of our method is illustrated in Fig. 1. First, a 3D face model is selected and decorated with accessorial models by automatic registration (Section 3). Then, the user's facial expressions are recognized by processing images captured from a webcam (Section 4). As a preprocessing for animating facial avatar, emotional templates and facial motion capture data are prepared for basic facial expressions. At runtime, these emotional data are used to augment an avatar's facial appearance and animate facial expressions in an automatic way (Section 5). Finally, the emotional avatar is rendered and animated on the screen according to the user's facial expressions (Section 6).

2 Related Work

Our work relates to the creation of facial avatars with details, and 3D facial animation based on facial expressions. Below we review the most relevant works with a focus on the approaches for unifying facial expression analysis and facial animation.

Facial Avatars with Details. With the advancement of computing and optical technologies, it is possible to acquire highly detailed 3D datasets of human faces. However, it is still very difficult to reconstruct facial features such as facial hair even using the latest optical equipment. To solve this problem, Beeler et al. [7]

∗ Corresponding author e-mail: smchoi@sejong.ac.kr

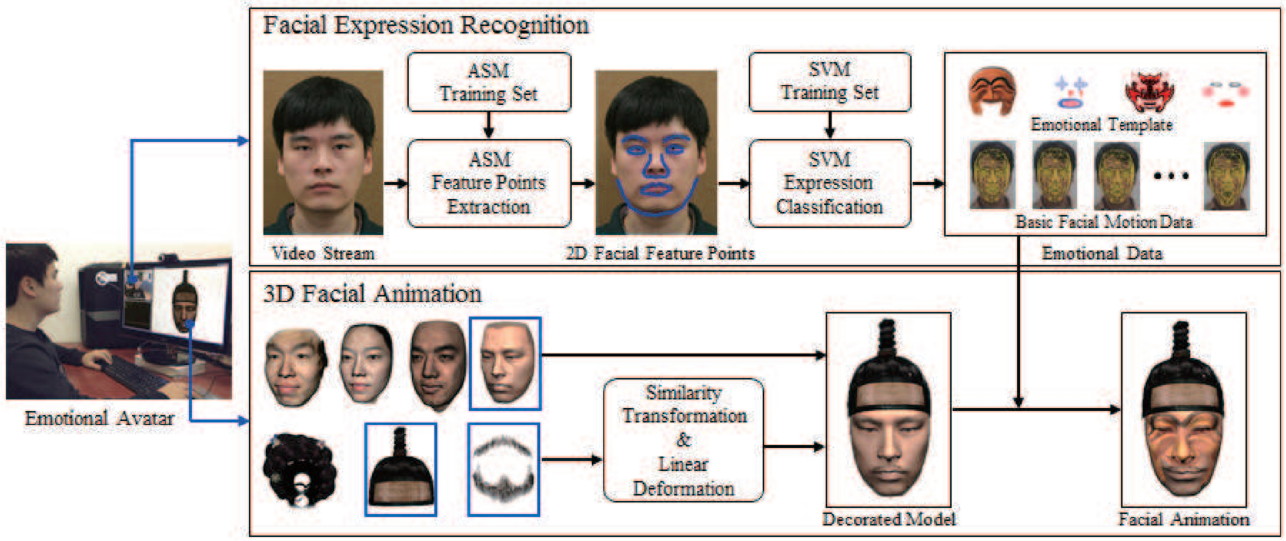

Fig. 1: An overview of the method used to create facial animation. After analyzing the user's facial expressions, the 3D facial avatar is animated according to the recognized expression with an emotional emphasis.

proposed a coupled 3D reconstruction method, which extracts facial hair from 2D images by a growing algorithm then reconstructs such facial hair by a multi-view stereo algorithm. Although this method shows realistic results for static faces, it is not suitable for facial animation because the method requires warping of a facial surface to combine facial hair with skin. As another approach, commercial graphics tools can be used to create 3D models of facial hair. However, this approach is limited to the predetermined character's face and cannot be easily applied to other faces without manual editing of the models. Instead of tedious editing, with our method, the user can easily apply accessorial models, such as face/head hair and hairstyles, to different types of faces by automatic registration.

Facial Expression Recognition. The study of facial expression analysis has been made mainly in the field of image-based computer vision. Most studies are based on a facial action coding system (FACS), which defines facial muscles in action units (AUs). Most facial expression analysis methods using FACS follow two main steps: facial representation for feature data acquisition and expression recognition by the analysis and classification of facial features. To obtain facial feature data, statistical models such as Active Appearance Models (AAMs) [8] and Active Shape Models (ASMs) [9] are mainly used. Support Vector Machines (SVMs) [10], a machine learning algorithm, is widely used for pattern analysis and classification of facial features. Recently, Shan et al. [11] presented an SVM-based method combined with Local Binary Pattern (LBP). Tsalakanidou et al. [12] also proposed an expression recognition method using AU rules which are made by computing vector differences between neutral and expressive states of a face. We employ ASMs for facial feature extraction and SVMs [13] for facial expression classification in same streamline, but our aim is to use the classified emotion to animate a 3D face with an emotional emphasis across facial expressions.

Facial Animation. There is vigorous research on combining facial expression analysis with facial animation. Stoiber et al. [14] presented an animation system that creates the human appearance space by analyzing a database of facial motion. The appearance space provides a coherent and continuous parameterization of human facial movements. With the advent of low-cost depth cameras, many methods using a depth camera have been developing to reconstruct 3D faces and estimate facial expressions. Weise et al. [15] developed a 3D avatar animation system that enables a user to animate the facial expression of the avatar using Kinect in real time. As another approach, Rhee et al. [16] presented a practical system for producing real-time facial expressions of a virtual avatar controlled by an actor's live performance using a single webcam. More recently, Cao et al. [17] introduced real-time animation system which can animate facial expressions only using a webcam without any depth camera. Although those systems can realistically animate facial expressions, their methods do not take into account factors beyond the realistic facial animation. Therefore, we aim to create an emotional avatar which can express emphasized emotions by combining realistic facial animation with facial expression analysis.

3 Automatic Registration of Accessorial Models with a Face Model

For the reconstruction of facial details or decoration purposes, accessorial models can be selected from the database to be attached to a face model. The selected 3D accessorial models are automatically registered with the facial surface using a deformation technique [18]. The main steps of our automatic registration algorithm are as follows. First, a set of feature points is determined from the vertices on an accessorial model. The feature points are classified into two groups: ST points for similarity transformation and LD points for linear deformation, as illustrated in Fig. 2. Second, similarity transformations are performed to reduce linear deformation errors. Then, the accessorial model is transformed using the angle of rotation, scale and distance parameters computed by the ST points. Finally, the accessorial model is linearly deformed to fit the surface of the face model. For the deformations, a set of target points are found, which are nearest to the LD points among the vertices on the face model, respectively. The other vertices on the accessorial model except the LD points are defined as free vertices.

For the linear deformations we employ the method [3], which demonstrates that the displacements of a set of sparse handle vertices provide sufficient geometric constraints for capturing the large-scale facial motion in a plausible way. Given as input constraints the 3D displacements $\mathbf{u}_H \in \mathbf{R}^{H\times 3}$ of the H handle vertices, the linear deformation $\mathbf{u}$ is computed by minimizing a simplified, quadratic thin shell energy. This amounts to solving the corresponding Euler-Lagrange equations

$$-k_s \triangle \mathbf{u} + k_b \triangle^2 \mathbf{u} = 0, \quad (1)$$

under the constraints $\mathbf{u}_H$ imposed by handle vertices. In this equation, k_s and k_b denote the stiffness for surface stretching and bending, respectively, and $\triangle$ is the Laplace-Beltrami operator. Discretizing the latter using the cotangent weights yields the linear system

$$\begin{pmatrix} \mathbf{A} & \mathbf{A}_H \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{u}_H \end{pmatrix} = 0, \quad (2)$$

to be solved for the unknown displacements $\mathbf{u}_H \in \mathbf{R}^{(V-H)\times 3}$ of the $(V-H)$ free vertices.

We pre-compute the “basis function” matrix $\mathbf{B} = -\mathbf{A}^{-1}\mathbf{A}_H$, according to the method presented by Bickel [18]. After factorizing $\mathbf{A}$ once, each column of $\mathbf{B}$ is computed by solving a sparse linear system involving $\mathbf{A}$ and the corresponding column of $\mathbf{A}_H$. At runtime, the linear deformation $\mathbf{u}$ is obtained from the handle displacement $\mathbf{u}_H$ by the matrix product $\mathbf{u} = \mathbf{B}\cdot\mathbf{u}_H$.

To naturally animate the combined accessorial model such as facial hair with the face model, it is necessary to control target points. The method for controlling target points is described in Section 5.

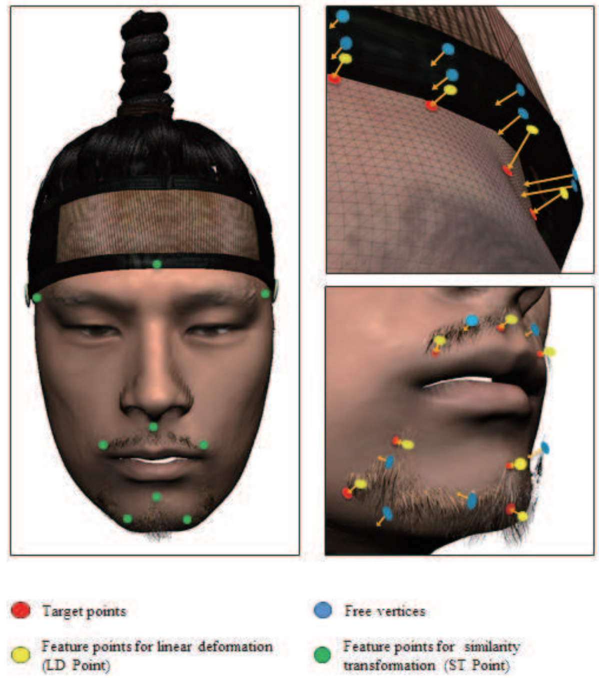

Fig. 2: Accessorial models such as facial hair or a topknot are linearly deformed on the surface of a 3D face model.

4 Facial Expression Recognition

For facial expression analysis, face location and major facial features such as eyes, eyebrows and mouth should be accurately extracted. Then facial movements should be encoded for facial expression recognition. For this purpose, we employ ASMs [19] for automatic extraction of facial features from input face images and SVMs [13] for the classification of the user’s facial expressions.

4.1 Extraction of Facial Features

To extract facial features using ASMs, it is needed to train various types of facial images. For the training, we construct a dataset that consists of sets of nine photos taken in different conditions for each subject. In our training images, the same landmark points as the dataset of DTU [20] are defined. The dataset used for facial image training is summarized in Table 1. To extract expressive facial features, we only use facial features mainly related to facial expressions in the ASM library, such as the eyes, eyebrows, nose, and mouth. The extracted feature points are used for classifying facial expressions and controlling expressive 3D facial animation.

Table 1: Image Dataset

Image Properties	Image Dataset Description
Image Type	Neutral(front, left, right) Happy, Angry, Surprise, Sadness, Neutral spot light added at the person's left side, Random expressions
Image Size	Width : 640 pixels, Height : 480 pixels
No. of Features	58 features
Gender	32 males, 11 females
Glasses	All images taken without glasses

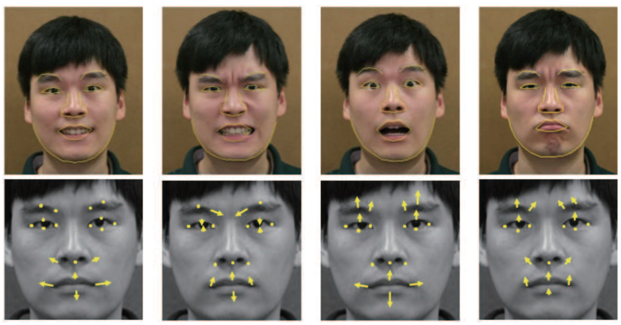

Fig. 3: Facial features from ASMs (top) and displacements of facial feature points from neutral to expressive states (bottom).

4.2 Classification of Facial Expressions

Facial classification using SVMs is divided into two stages: the stage of generating a set by training the user's facial expressions and the stage of facial feature classification using the generated training set when new input facial expressions are entered. The inputs for generating the training set are distance vectors calculated by feature points extracted from ASMs, according to the AU codes of FACS. For example, the distance vectors from the eyebrows to the eyes are used for AU codes 1, 2, and 4. In Fig. 3, the extracted feature points by ASMs are shown in yellow at the top, and the resulting distance vectors are illustrated in yellow at the bottom, which illustrates the differences between neutral and expressive face images.

In order to classify the facial expression of the input video clip, we first generate an SVM training set for each of the following five expressions: neutral, happy, angry, surprised, and sadness. When a new input is entered from the webcam, the facial expression from each SVM training set is examined then the SVM training set with the highest weight is assigned to the user's current expression. To determine which SVM training set has the highest weight, we use AUs for facial expressions in the Cohn-Kanade database [21].

5 Animation of Emotional Avatars

In this section, we first describe how to animate a facial avatar using a set of displacements of feature points over time for each facial expression. The displacements of feature points for basic facial expressions are acquired by a widely used RGB-D camera (see Fig. 1) as a preprocessing step. Second, we explain how to apply the facial features from ASMs in Section 4 to the displacements of the basic facial motion data, which is captured by using Candide-3's feature points [22]. Finally, we show how the classified facial expression using SVMs can be combined with emotional templates to emphasize the input facial expression from a webcam in real time.

5.1 Generating Displacement Sets of Feature Points

The displacements of feature points can be acquired by sequentially tracking the changing images from neutral to expressive faces using a RGB-D camera and a face tracking API to detect AU codes [22]. To generate a set of displacements, motion capture data representing facial movements should be prepared. The motion capture data is defined as $M_{i,t} \in \mathbf{R}^3, i = \{1,...,97\}, t = \{0,1,...\}$. M is a predefined handle vertices based on Candide-3 and t is a facial image over time. The set of displacements is defined as $D_{i,t} = (M_{i,t} - M_{i,0})$.

We classify facial expressions into four categories except neutral, and then generate four sets of displacements from neutral to different expressive states to animate facial expressions. To animate facial expressions using D, the handle vertices $H_{i,0}$ are also defined on the face model according to the same position as $D_{i,0}$. D is similarity transformed to D' to match the size of D with $H_{i,0}$.

To animate a 3D face model over time, we also use a linear deformation mentioned in Section 3. As illustrated in Fig. 2, the free vertex is defined as a vertex that does not belong to $M_{i,0}$ among the vertices on the face model. The 3D face model is sequentially deformed according to the displacements D' at each frame. Not only the feature points in M are deformed, but the free vertices are also moved by using the displacements D' and predetermined weights.

This linear deformation for animation is only applied to the face model; hence accessorial models should search for target points repeatedly (see Fig. 2). During such deformations, there can be gaps between the face model and accessorial models otherwise some protrusion may occur because the accessorial models are deformed after being modified by initial similarity transformation. Since the initially specified target point is identical with a vertex on the 3D face model, we save the index number of the vertex and fix the next target point with the saved index number to avoid this distortion during deformation.

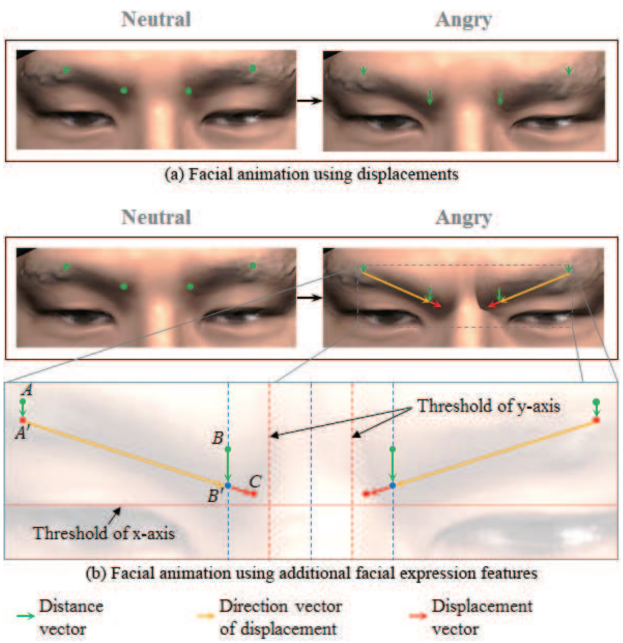

Fig. 4: Animation of a 3D facial avatar according to the presence or absence of additional facial expression features from the user.

5.2 Using Information from the User's Facial Expressions

The face tracking APIs capture some movements related to AU codes, so that they capture only specific movements. For example, the expression of anger characterizes both sides of the inner eyebrows narrowed to appear as wrinkles between the eyes [23]. However, it is difficult for general face tracking APIs to capture the detailed movement of the inner eyebrows narrowed. The API using Candide-3's feature points [22] represents an angry expression by distance vectors in the eyebrows illustrated by the green arrows in Fig. 4 (a).

To animate natural expressions, we compute more detailed facial movements by combining such distance vectors with the facial features from ASMs (see Fig. 4 (b)). As mentioned in Section 4, this additional information from ASMs can be acquired when the user expresses his/her emotions in front of a webcam in real time. To compute the detailed movements for narrowed inner eyebrows, the distance vectors are used as weights for moving from point B' to point C, since the movement towards the glabella and downward eyebrows occurs simultaneously. The thresholds of weights are shown in the bottom of Fig. 4 (b). The direction of displacement vectors is defined as $\overrightarrow{A'B'}$.

The expression of happiness can be represented as a state with a closed mouth or visible teeth in a smile. The lip movements can be obtained by using face tracking APIs. Similar to eyebrows, the detailed movements for lips are also computed by additional facial features from ASMs. To control the shape and size of an opened mouth, the weights can be computed by using the distance vector between the middle point of the upper lip contour and the middle point of the lower lip contour. Hence, the user-specific expressive animation can be achieved by applying such detailed movements of the eyebrows and lips to D'.

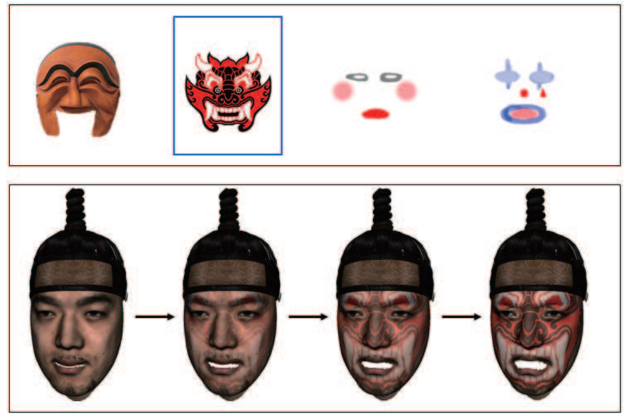

Fig. 5: Emotional templates (top) and the creation of intermediate emotional states from neutral to angry (bottom).

5.3 Applying Emotional Templates to a 3D Face

To create an emphasized emotional face, we use artistic emotional templates, as shown in the top of Fig. 5. As an example, we use traditional masks for performances and facial paintings for special characters as emotional templates. Besides such templates, users can freely create their own. Then, the user can choose four emotional templates representing the classified four facial expressions from the template database. Each emotional template is aligned with a 3D face model using Candide-3's feature points [22]. At runtime, the texture of the emotional template can be easily combined with the skin texture of the 3D face model in its neutral state showing different moods. Hence, just by expressing his/her emotions, the user can change the emotional mood of the facial avatar and easily synthesize the intermediate states between the different emotions in an artistic way. In Fig. 5, the bottom shows the interpolated emotional avatar according to the change from neutral to angry.

6 Experimental Results

The proposed animation system was implemented on a PC with an Intel Core i7-3770 CPU and a webcam with

Fig. 6: The user interface unifies facial expression analysis and facial animation.

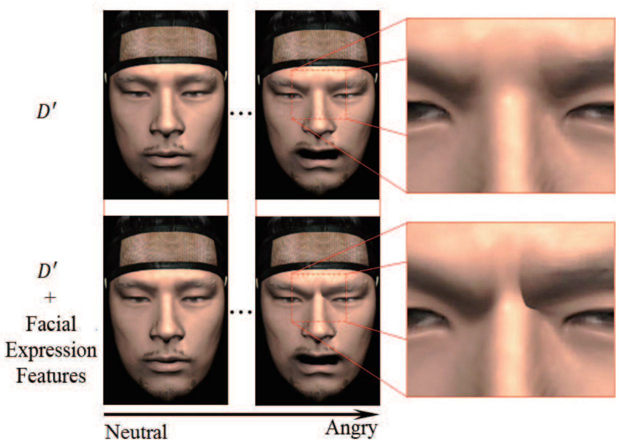

Fig. 7: A comparison result according to additional facial expression features.

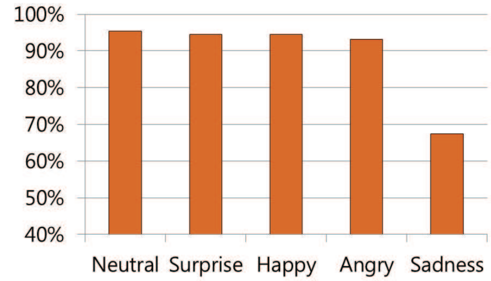

Fig. 8: Facial expression recognition rates(%) under the per frame classification scenario.

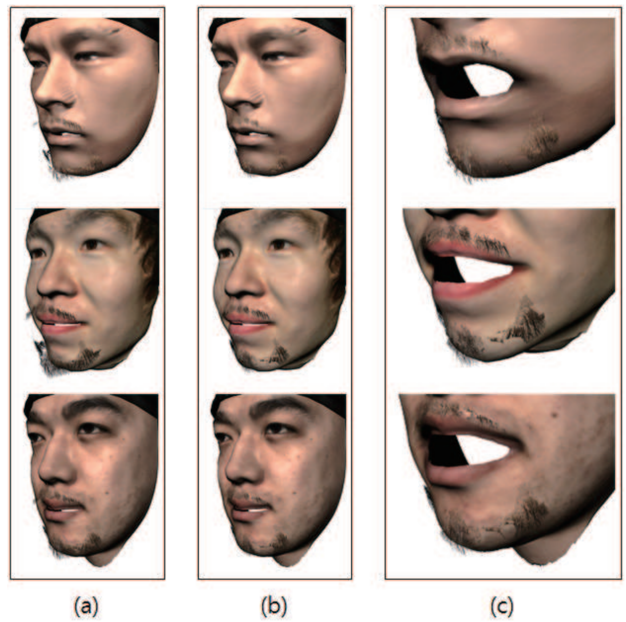

Fig. 9: The automatic registration of facial hair with a male face: initial state (a), final results (b), enlarged results (c).

640 x 480 video recording at 30fps, as shown in Fig. 6. This system allows a combination of a 3D face model with accessorial models (left) to offer an animatable decorated facial avatar. It also extracts the user's facial features in the video stream from the webcam in real time (top right), and recognizes the user's facial expression (bottom right). The 3D emotional avatar is animated according to the user's facial expressions (bottom left).

Fig. 7 shows the result of the comparison between animation using D' (top) and animation using D' and the user's facial expression features (bottom). As you can see, the animation result using additional facial expression features is more expressive and characteristic (See the red box in Fig. 7).

Fig. 8 shows the facial expression recognition rates for animation control. We can see a high percentage of prediction accuracy for each expression except sadness. In our study, sadness shows low accuracy because the displacements between sad and neutral expressions are relatively small.

For the animation of male avatars with facial hair, the facial hair around the mouth should be accurately registered because there are a lot of movements especially around the mouth. Fig. 9 shows the results of animation of male avatars with facial hair. Fig. 9 (a) shows the unregistered initial state of a facial model and facial hair, (b) shows the results of automatic registration, and (c) enlarges the images around of the mouth. We also show emotional male and female avatars animating according to the change of the user's facial expressions in Fig. 10.

Fig. 10: From left to right: a decorated 3D face in a neutral state, happy, angry, surprised, and sadness.

7 Conclusion

We presented an emotional avatar to express the user's emphasized emotions in an artistic way. The method achieves not only visual realism with some details such as facial hair and special hairstyles, but it also provides behavioral realism by incorporating the user's notable facial expressions with the basic facial motion data. The proposed emotional avatar is also useful to keep people anonymous by hiding their faces in the cyberspace using diverse facial painting. Our system can be used for interesting applications such as decorating avatars according to personal taste, design and simulation for special makeup, etc. In particular, the incorporation of vision-based control and animation of facial avatars with a decoration function can be used as a noble interface for many human-computer interaction applications such as tele-presence systems, intelligent dialogue systems, etc.

Acknowledgement

This research was supported by Ministry of Culture, Sports and Tourism and KOCCA in the CT R&D Program 2011. Also, this work was partially supported by BRL program through the NRF of Korea (2010-0019373).

References

[1] R.W. Picard, Affective Computing, MIT Press, 2000.
[2] S.-M. Choi and Y.-G. Kim, "An affective user inferface based on facial expression recognition and eye-gaze tracking", Affective Computing and Intelligent Interaction, LNCS, **3784**, pp.907-914, 2005.
[3] B. Bickel, M.Botsch, R. Angst, W. Matusik, M.A. Otaduty, H. Pfister and M. Gross, "Multi-scale capture of facial geometry and motion", ACM Trans. Graph, **26**, 33:1-33:10, 2007.
[4] T. Beeler, B. Bickel, P. Beardsley, R. Sumner and M. Gross, "High-quality single-shot capture of facial geometry", ACM Trans. Graph, **29**, 40:1-40:9, July 2010.
[5] D. Bradley, W. Heidrich, T. Popa and A. Sheffer, "High resolution passive facial performance capture", ACM Trans. Graph, **29**, 41:1-41:10, July 2010.
[6] T. Beeler, F. Hahn, D. Bradley, B. Bickel, P. Beardsley, C. Gotsman and M. Gross, "High-quality passive facial performance capture using anchor frames", ACM Trans. Graph, **30**, 75:1-75:10, Aug 2011.
[7] T. Beeler, B. Bickel, G. Noris, P. Beardsley, S. Marschner, R. W. Sumner and M. Gross, "Coupled 3D Reconstruction of Sparse Facial Hair and Skin", ACM Trans. Graph, **31**, 117:1-117:10, July 2012.
[8] T. Cootes, G. Edwards and C. Taylor, "Active appearance models", IEEE Transactions on Pattern Analysis and Machine Intelligence, **23**, 681-685, June 2001.
[9] T. Cootes, "An introduction to active shape models", Image Processing and Analysis, pp.223-248, Oxford University Press, 2000.
[10] I. Kotsia and I. Pitas, "Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines", IEEE Trans. on Image Processing, **16**, 172-187, January 2007.
[11] C. Shan, S. Gong and P. W. McOwan, "Facial expression recognition based on local binary patterns", Image and Vision Computing, **27**, 803-816, 2009.
[12] F. Tsalakanidou and S. Malassiotis, "Real-time 2D+3D facial action and expression recognition", Pattern Recognition, **43**, 1763-1775, 2010.
[13] C.-C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines", ACM Trans. Intell. Syst. Technol. **2**, 27:1-27:27, April 2011.
[14] N. Stoiber, R. Seguier and G. Breton, "Facial animation retargeting and control based on a human appearance space", Computer Animation and Virtual Worlds, **21**, 39-54, January 2010.
[15] T. Weise, S. Bouaziz, H. Li and M. Pauly, "Realtime performance-based facial animation", ACM Trans. Graph, **30**, pp.77:1 -77:10, July 2011.
[16] T.Rhee, Y. Hwang, J.D. Kim, C. Kim, "Real-time facial animation from live video tracking", In Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 215-224, August, 2011.
[17] C. Cao, Y. Weng, S. Lin and K. Zhou, "3D shape regression for realtime facial animation", ACM Trans. Graph, **32**, 44:1-44:10, July 2013.
[18] B. Bickel, M. Lang, M. Botsch, M. Otaduy and M. Gross, "Pose-space animation and transfer of facial details", In Proceedings of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 57-66, July 2008.
[19] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model", In Proceedings of the 10th European Conference on Computer Vision (ECCV), pp.504-513, 2008.
[20] M. B. Stegmann, B. K. Ersboll and R. Larsen, "FAME-a flexible appearance modelling environment", IEEE Trans. on Medical Imaging, **22**, pp. 1319-1331, 2003.
[21] P. Lucey, F. J. Cohn, T. Kanade, J. Sarahih and Z. Ambadar, "The extended Cohn-Kanade dataset(CK+): A complete dataset for action unit and emotion-specified expression", Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 94-101, 2010.
[22] J. Ahlberg, "CANDIDE-3: an updated parameterised face", Dept. of Electrical Engineering, Linkping University, January 2001.
[23] P. Ekman, W.V. Friesen and J. Hager. "Facial action coding system: research nexus", Network Research Information, Salt Lake City, UT, USA, 2002.

Taehoon Cho is received his M.S. degree in Computer Science and Engineering from Sejong University at Seoul, Korea in 2014. His research interests include 3D facial animation, 3D motion capture, 3D printer, human-compupter interaction, image processing and computer graphics.

Jin-Ho Choi is received M.S. degree in Computer Science and Engineering from Sejong University at Seoul, Korea in 2014. Currently he is a developer in 10DR of the Department of Solution Development. His research interests include medical image processing and computer graphics.

Hyeon-Joong Kim received his Ph.D. degree in Computer Science and Engineering from Sejong University at Seoul, Korea in 2014. Currently he is a software engineer in 3D Systems Korea. His research interests include computer graphics, virtual reality, point-based graphics, realistic facial rendering, and 3D printing.

Soo-Mi Choi is a Full Professor in the Department of Computer Science and Engineering at Sejong University, Seoul, Korea. She received her B.S., M.S. and Ph.D. degrees in Computer Science and Engineering from Ewha University of Korea in 1993, 1995 and 2001, respectively. After she received her Ph.D., she joined the Center for Computer Graphics and Virtual Reality (CCGVR) at Ewha University as a Research Professor. In 2002, she became a faculty member in the Department of Computer Engineering at Sejong University. From September 2008 to August 2009 she was a Visiting Scholar at the CG Lab of ETH Zurich in Switzerland. Her current research interests include computer graphics, virtual reality, human-computer interaction, and ubiquitous computing.