Exaggeration of facial expressions from facial motion capture data

Seongah Chin (陈诚儿)* **and Chung-Yeon Lee (李充然)**

Division of Multimedia, College of Engineering, Sungkyul University,
#147-2 Anyang-8 Dong. Manan-Gu, Anyang City, 430-742, Korea
**E-mail: solideo@sungkyul.edu*

Received June 17, 2009

We propose a method for exaggerating facial expressions derived from exaggeration mappings that transform facial motions into exaggerated motions. The exaggeration mapping of facial motions is defined by non-negative matrix factorization. Three-dimensional facial expressions are simulated using the exaggerating rate as an input value to convey the degree of variation. Experiments show the validity of the exaggeration mapping and facial expression simulations.

OCIS codes: 100.0100, 110.0110.
doi: 10.3788/COL20100801.0029.

Facial expressions have been widely used for characters, digital actors, and virtual humans in various fields such as video games, movies, social agents, commercials, and anthropology. The human face, in particular, plays a significant role in non-verbal communication, with a plethora of expressions that vary by personality. Most researchers in the area of facial expression synthesis seem to focus on realistic expressions or retargeting[1−3], rather than exaggerating expressions. To draw attention to subjects, it is very important to exaggerate facial expressions, but this is a labor-intensive process. As a consequence, a method for automatic exaggerated facial expression synthesis is a must. Exaggeration is considered to be similar to placing more emphasis on the key features of an object. It should be noted that the larger accentuated features are found before the lesser features are determined. However to our knowledge, no published articles have reported on the exaggeration of facial expressions. In this letter, we propose a method to exaggerate facial expressions derived from facial motions acquired by motion capture. In this transformation, we employ non-negative matrix factorization (NMF) to realize exaggeration mapping. The proposed method is validated by showing some experimental results.

Exaggeration mapping is used to transform a facial motion composed of feature points on a face acquired by motion capture into an exaggerated facial motion. The fundamental idea of this approach to the exaggeration mapping of facial motions is to place more emphasis on those features with relatively more movements. The first step is proper data factorization. NMF is suitable for parts-based data factorization because of its advantages over principal component analysis (PCA) or vector quantization (VQ) with respect to its use of non-negative constraints. In addition, NMF learns parts of faces, whereas PCA and VQ learn holistically. These constraints lead to a parts-based representation computed only by additive, rather than subtractive, combinations[4,5].

The mapping method primarily comprises two procedures. Firstly, NMF decomposes a facial motion matrix into a basis and its weight matrix. Then, a facial motion is exaggerated by multiplying both the weights and residuals acquired from the NMF decomposition by a specific exaggeration rate.

Cross-cultural research presents six universal expressions — surprise, fear, disgust, anger, happiness, and sadness — for which a sequence of facial markers, called a facial motion, is factorized in order to exaggerate individual facial markers of the facial motion. Each motion consists of the three-dimensional (3D) movements of markers attached to the principal facial muscles of an actor. These facial motions are regarded as an $n \times m$ matrix M, each column of which consists of the x, y, and z coordinates of the markers. Given a facial motion denoted by M, NMF decomposes M into two matrices B and E, as

$$M_{i\mu} \approx (BE)_{i\mu} = \sum\nolimits_{a=1}^{r} B_{ia} E_{a\mu}, \tag{1}$$

to approximate the facial motions. The dimensions of the factorized matrices B and E are $n \times r$ and $r \times m$, respectively, with r satisfying $(n + m)r < nm$.

Each column of the matrix B contains a basis vector, while each column of E includes the weights corresponding to the measurement column in M using the basis from B. To estimate the factorization matrices, an objective function has to be defined. This objective function works out the likelihood of computing the facial motions in M from the basis B and encodings E. The objective function that we used is given as

$$H = \sum\nolimits_{i=1}^{n} \sum\nolimits_{\mu=1}^{m} [M_{i\mu} \log(BE)_{i\mu} - (BE)_{i\mu}]. \tag{2}$$

Solutions for NMF begin with initializing the non-negative conditions for B and E. Continuing the iteration of the update rules in

$$B_{ia} \leftarrow B_{ia} \sum\nolimits_{\mu} \frac{M_{i\mu}}{(BE)_{i\mu}} E_{a\mu}, \tag{3}$$

$$E_{a\mu} \leftarrow E_{a\mu} \sum\nolimits_{i} B_{ia} \frac{M_{i\mu}}{(BE)_{i\mu}}, \tag{4}$$

M finds an approximate factorization $M \approx BE$ by converging to a local maximum of the objective function

1671-7694/2010/010029-04

given in Eq. (2).

A fundamental idea of exaggerating facial motions is an exaggeration of the differences from the mean, emphasizing the movements of the distinctive feature points of a facial motion. Provided that the objective function stops iterating after E is properly acquired, E needs to be divided again into the mean m_i and deviation d_i of each column of E for its exaggeration. Each dimension is composed of a basis vector $\boldsymbol{b}_i$ (the ith column of the matrix B) and its weights, including expectation m_i and deviation d_i of the ith row in the matrix E. The residual r is added to eliminate noise that has been made in the process of the matrix factorization. The facial motion

$$\boldsymbol{f} = \sum\nolimits_i e_i \cdot \boldsymbol{b}_i + r = \sum\nolimits_i (m_i + d_i) \cdot \boldsymbol{b}_i + r \qquad (5)$$

is in the form of a non-negative linear combination of the basis and a residual.

The exaggerated facial motion $\boldsymbol{f}'$, called exaggeration mapping, is calculated by scaling the deviation d_i and residual r with an exaggeration rate Θ :

$$\boldsymbol{f}' = \sum\nolimits_i (m_i + t \cdot d_i) \cdot \boldsymbol{b}_i + \Theta \cdot r, \qquad (6)$$

where $t = 1$ if $|d_i| < 2 \cdot s_i$ and $t = \Theta$ if $|d_i| \geq 2 \cdot s_i$ with $|d_i| = |e_i - m_i|$ and standard deviation s_i. Figure 1 shows the comparisons between original surprise motion and exaggerated surprise motions with exaggeration rate $\Theta = 1.0$ (original), 1.2, and 1.5, which distinctively displays high emphasis on the plot of $\Theta = 1.5$.

In addition, it shows consistent curve patterns, even though changing the exaggeration rate Θ=1.0, 1.2, or 1.5 transforms the values. The peak is also denoted by a circle. We plotted three change rate patterns calculated by the distance between the neutral frame and each anger motion frame, as shown in Fig. 2. The distinctions are clearly detected in the plot, showing that the differences vary as the exaggeration rate increases. The vertical line indicates a high peak frame. The change rate becomes 164.02, the highest value for surprise motion, at the 170th frame. The other two values are 109.37 and 131.22, corresponding to $\Theta = 1.0$ and 1.2 exaggeration rates, respectively, as also shown in Fig. 2.

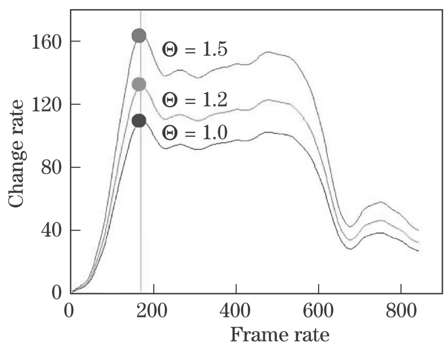


Fig. 2. Curves of change rates computed by the distance between the neutral frame and frames for anger motion with exaggeration rates $\Theta = 1.0$, 1.2, and 1.5.

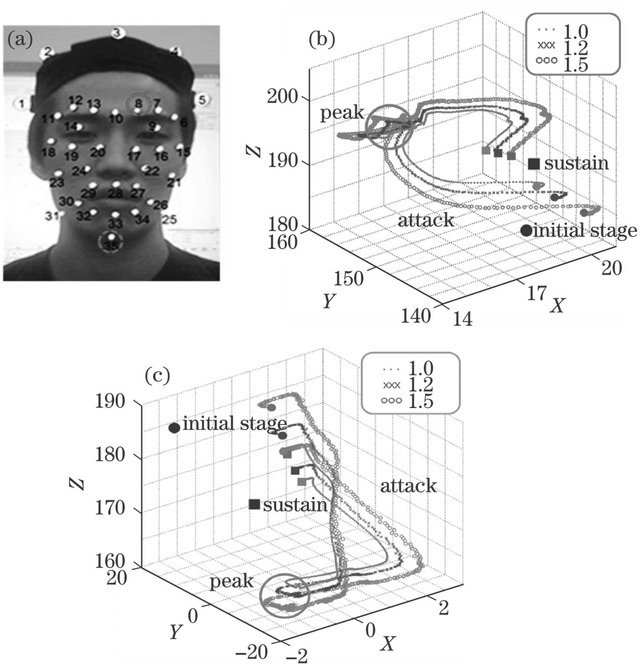


Fig. 1. (a) Facial marker positions, (b) curve patterns of the 8th marker, and (c) curve patterns of the 35th marker for surprise motions with exaggeration rates Θ = 1.0, 1.2, and 1.5.

The experimental results can be summarized with respect to three approaches, including facial motion capture, validation of the exaggeration mapping, and exaggerated facial expressions.

We asked an actor to visualize six universal expressions as realistically as possible, taking into consideration three expression phases (i.e., an initial attack phase, a sustaining phase, and a relaxation phase) in order to properly capture facial motions and analyze them. The positions of the markers were determined by carefully considering the muscle movements.

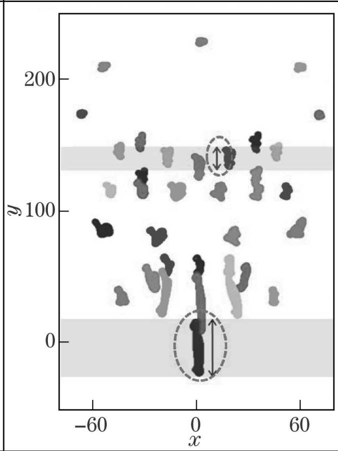
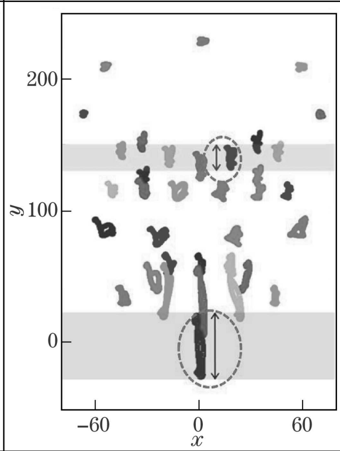
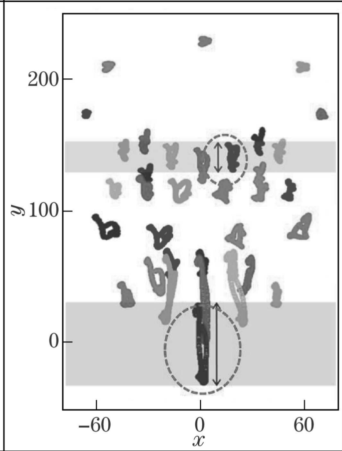
To validate the exaggeration mapping, we show anger motion composed of a sequence of 35 marker locations with the performance time periods shown in Table 1. The 8th and 35th markers' motions are indicated in dotted circles. Their maximum distances become distinguishably larger as the exaggeration rates, Θ = 1.0, 1.2, and 1.5, increase. This observation implies that exaggeration mapping is considerably acceptable since the facial motions are exaggerated by exaggeration rates.

To create exaggerated facial expressions, we employ a muscle-based method[6] that uses 18 linear muscles whose contraction and expansion movements can be explained by a fan-shaped geometry. Three sphincter muscles are also used to express eye blinking and mouth movement with ellipsoidal principles. However, the eye does not have an ideal elliptical shape. Accordingly, some variation in terms of distortion of the eye shape should be allowed. In previous research, the authors developed customized motion cloning to achieve facial expressions for general users. The muscle structure and a more detailed description can be found in the previous article[7].

Figure 3(a) shows surprise, happiness, sadness, anger, fear, and disgust expressions, demonstrating the facial motions when the exaggeration rate Θ is set to 1.0, 1.2, and 1.5. It can be seen that the marker movements vary with the changes in the exaggeration rate. An additional experiment with a different face model is shown in Fig. 3(b) to demonstrate the validity for general users.

As previously noted, exaggerating the facial motions is

Table 1. Validation of Exaggeration Mapping

	$\Theta = 1.0$	$\Theta = 1.2$	$\Theta = 1.5$
Anger			
Distance of the 8th Marker	15.2492	18.3748	22.9685
Distance of the 35th Marker	51.6178	61.9275	77.4093

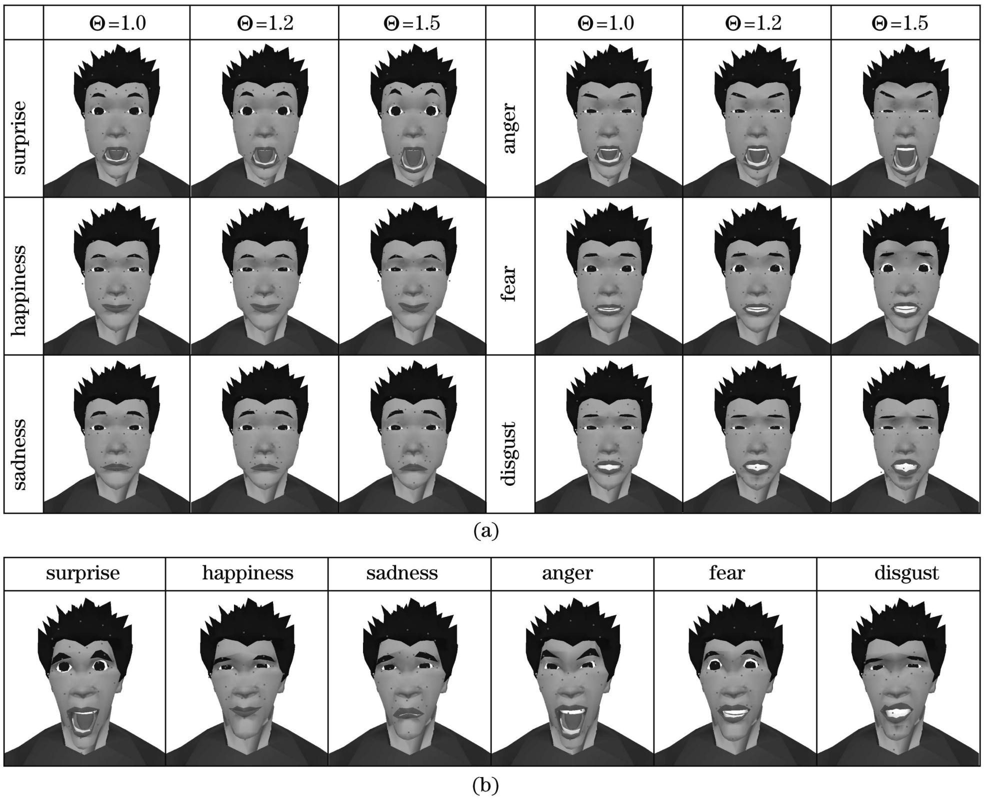


Fig. 3. (a) Six expressions displayed by facial motions (markers) corresponding to exaggeration rates $\Theta = 1.0$, 1.2, and 1.5; (b) six expressions for a different face model with exaggeration rate $\Theta = 1.5$.

the first step toward the creation of exaggerated facial expressions. Based on the exaggerated facial motions, we can manipulate the muscles by selecting proper contraction values to reach the ultimate facial expressions. Empirical experiments were used to determine the 1.0 – 1.5 range for the exaggeration rates. Finally, the exaggerated facial expressions are simulated by manipulating the muscles.

In conclusion, we present a method focusing on a technique for the automatic exaggeration of facial expressions. We employed NMF is used to decompose the facial motion acquired by motion capture into a suitable form for transformation. We analyze the facial motions and implement facial expressions to show the validity of the proposed method.

This work was partially supported by the Korea Research Foundation Grant funded by the Korea Government under Grant No. KRF-2008-521-D00398.

References

1. Q. Zhang, Z. Liu, B. Guo, D. Terzopoulos, and H. Shum, IEEE Trans. on Visualiz. Computer Graph. **12,** 48 (2006).
2. I. Kotsia and I. Pitas, IEEE Trans. Image Process. **16,** 172 (2007).
3. Y. Wang, X. Huang, C. Lee, S. Zhang, and Z. Li, Computer Graph. Forum **23,** 677 (2004).
4. D. D. Lee and H. S. Seung, Nature **401,** 788 (1999).
5. P. O. Hoyer, J. Mach. Learning Res. **5,** 1457 (2004).
6. F. I. Parke and K. Waters, *Computer Facial Animation* (A. K. Peters, Wellesley, 1996).
7. S. Chin and K. Y. Kim, IEEE Trans. System Man Cybern. C **39,** 315 (2009).