H-036

Subtle Facial Expression Recognition Based on Expression Category-dependent Motion Magnification

Lumei Su[1] Shiro Kumano[2] Kazuhiro Otsuka[2] Dan Mikami[2]
Junji Yamato[2] Eisaku Maeda[2] Yoichi Sato[1]

Abstract

Previous subtle facial expression recognition studies have primarily focused on extracting the subtle changes of facial expression but not classifying the extracted subtle expression into categories such as anger and happiness. In this paper, we propose a novel method based on expression category-dependent motion magnification to recognize subtle facial expressions. Because it's difficult to discriminate subtle expressions directly from raw subtle facial deformation data, we use expression category-dependent motion magnification to reliably transform subtle facial expressions into the corresponding exaggerated facial expressions. The magnified facial expressions become much more discriminable and we can infer the expression category of the subtle expression from them. Experiments using our database showed that the proposed method has a promising performance on subtle expression recognition, especially in the early stage of making facial expression.

1. Introduction

Facial expression is one of the most important ways that people communicate emotion and other mental signal besides verbal expressions. As an active and challenging research topic in computer vision, automatic facial expression recognition impacts important applications in areas such as human-computer interaction and human affective recognition.

Most facial expression recognition methods focus only on extreme facial expressions. However, in real-world environment, people express facial expressions with different magnitude, such as subtle expressions. Therefore, recognizing subtle facial expressions is desirable and important for natural human computer interaction.

Several researchers have attempted to develop techniques to recognize subtle facial expressions [1-5]. Most of them have focused on representing the subtle facial changes in facial expressions but not classifying the extracted subtle expression. Song et al. [3] proposed a method based on vector field decomposition to model subtle facial expressions, but they did not attempt to recognize the extracted subtle expression. Others have paid attention to recognizing subtle facial expressions [1, 5]. However, there are still few effective methods for classifying subtle facial expressions. Sun et al.[5] integrated a 3D facial surface descriptor and 2D hidden Markov models (HMMs) to recognize the expression category of low intensity expression sequences (sequences close to the starting or ending frames showing near-neutral expressions), but they did not describe clearly that how low intensity expression sequences they recognized. Park et al.[1] proposed an simple motion magnification method to recognize subtle facial expressions by multiplying the extracted subtle facial motion vectors. However, Park et al. fixed magnification factors decided empirically. It may take time to choose appropriate magnification factors. Moreover, their motion magnification didn't consider different expression category cases of pre-magnified subtle expression. Therefore, they may not guarantee to transform subtle facial expressions into the corresponding exaggerated expressions by using the fixed magnification factors.

In this paper, we propose a novel expression category-dependent motion magnification method to recognize subtle facial expressions. We use a 3D motion capture system to accurately capture subtle facial deformation that occurs when forming facial expressions while most of existed methods are based on 2D facial expressions. Then we magnify the captured subtle expression into the corresponding exaggerated expression based on expression category- and intensity-dependent magnification factors. Therefore, the magnified expression becomes much more discriminable. Finally, we infer the subtle expression category from the magnified expressions.

The novelty of this work is that we propose an effective method to learn appropriate and variable magnification factors for pre-magnified subtle expressions. We consider sufficient possible cases of pre-magnified subtle expressions (the expression category and the degree of change of pre-magnified facial expressions), and for each pre-magnified subtle expression case, we prepare an expression category-dependent and intensity-dependent magnification factor from the training expression database. Therefore, based on the learned magnification factors, we can reliably transform subtle facial expressions into the corresponding exaggerated expressions, leading to a significant improvement in the performance of subtle expression recognition.

2. Subtle facial expression recognition based on motion magnification

Figure 1 shows the overall framework of the proposed method, which consists of two stages: training and testing. In the training stage, we first learn an exaggerated facial expression template with apex deformation for each expression category from the training database. After that, we learn expression category-dependent and intensity-dependent motion magnification factors by using the learned exaggerated expression templates. In this paper, the intensity of expression means the degree of facial expression deformation. For a given subtle facial expression in certain category with certain intensity, we can use a category-dependent and intensity-dependent magnification factor to tra-

1 Tokyo University
2 Nippon Telegraph and Telephone Corporation

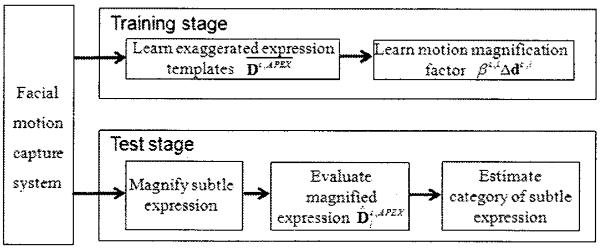

Figure 1. Flow chart of proposed system

nsform this subtle expression into its corresponding exaggerated one, and the intensity of the magnified expression can reach apex expression deformation. Therefore, each learned magnification factor corresponds to certain expression category and intensity hypothesis of pre-magnified expression. In the test stage, we firstly search all possible magnified expressions by magnifying the test facial expression with all learned magnification factors. The magnified expression with the best result is most similar to the learned corresponding exaggerated expression template. Then, we evaluate each magnification result by computing the distance between the magnification result and its corresponding exaggerated expression template. After that, by choosing the category-dependent and intensity-dependent magnification factor that generates the magnified expression with the best result, we can infer the category of test subtle expression.

2.1 Facial motion model

To model facial motion, we collected expression sequence data starting from a neutral expression, and use the concatenated displacement of all 3D facial markers at each frame t to describe facial motion feature, which is denoted as $\mathbf{D}_t \in R^{3\times M}$. The displacement of the j_{th} marker at each frame t is specified as $\mathbf{d}_{t,j} = [d_{j,x}, d_{j,y}, d_{j,z}]$. Figure 2 shows the distribution of $M = 26$ facial markers on the human face, which is denoted as j, where $j = 1, 2, ..., M$. The M facial markers' positions were determined by carefully considering muscle movements. In this paper, we use facial motion feature to represent the facial deformation that occurs when forming facial expressions

We utilize a high-speed, maker-based optical facial motion capture system to capture 3D coordinate value of M facial markers when people is expressing facial expression. The speed of the motion capture system reaches 100 frames per second, which enables it to capture subtle facial motion.

After that, we need to align the raw 3D data from the world coordinate system into a facial coordinate system, which is shown in Fig.2. First, global transformation is removed from the measured 3D coordinates by using four markers rigidly fixed to a user's head. Resulting data are given in the local coordinate system defined on a face using several markers including the one on a nose and two markers on eyelids. The coordinate value of all facial markers at each frame t in facial coordinate system is represented as $\mathbf{x}_t = [x_1, y_1, z_1, \dots, x_M, y_M, z_M]$. One collected expression sequence is specified by a set of $X = \{\mathbf{x}_0, ..., \mathbf{x}_N\}$, where N denotes the total number of frames in the expression sequence. Therefore, the displacement of all 3D facial markers at each frame $\mathbf{D}_t$ is obtained from $\mathbf{D}_t = \mathbf{x}_t - \mathbf{x}_0$, where $\mathbf{x}_0$ is the coordinate value of all

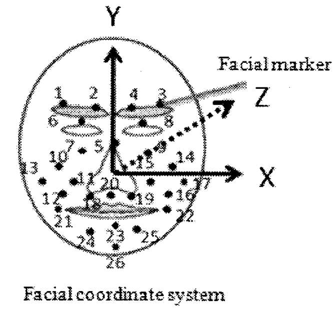

Figure 2. Distribution of 26 facial markers

facial markers at first frame (neutral expression).

2.2 Training stage

In this paper, motion magnification means magnifying the facial deformation that occurs when forming facial expressions and we use the concatenated displacement of all facial markers $\mathbf{D}_t$ to represent facial deformation. Facial deformation is governed by two factors, the category of expression and the magnitude of facial deformation [7]. Therefore, by using a proper motion magnification considering the expression category c and intensity i of the pre-magnified subtle expression $\mathbf{D}_t^{c,i}$, we can reliably magnify subtle facial expression $\mathbf{D}_t^{c,i}$ into its corresponding exaggerated facial expression $\mathbf{D}^{c,APEX}$, where expression category $c \in \{\text{Happiness, Anger, Surprise, Fear, Sadness, Disgust}\}$, intensity $i \in \{1, 2, ..., 9, 10\}$ denotes the degree of facial deformation causing by facial expression, and $\mathbf{D}^{c,APEX}$ denotes exaggerated facial expression with expression category c and apex deformation (intensity $i = 10$). In our approach, the motion magnification is implemented on each facial marker individually. A proper motion magnification for the j_{th} facial marker should satisfy:

$$\mathbf{d}_j^{c,APEX} = \mathbf{d}_{t,j}^{c,i} + \beta_j^{c,i}\Delta\mathbf{d}_j^{c,i} \tag{1}$$

where $\mathbf{d}_j^{c,APEX}$ is the displacement of the j_{th} facial marker for exaggerated expression with category c and apex deformation, $\mathbf{d}_{t,j}^{c,i}$ is the displacement of the j_{th} facial marker for subtle expression with category c and intensity i, $\beta_j^{c,i}$ is the magnification scale for $\mathbf{d}_{t,j}^{c,i}$, $\Delta\mathbf{d}_j^{c,i}$ is the magnification direction for $\mathbf{d}_{t,j}^{c,i}$, and $\|\Delta\mathbf{d}_j^{c,i}\| = 1$. Figure 3 illustrates motion magnification of the j_{th} facial marker from subtle *Surprise* expression with intensity ($i = 2$) to apex *Surprise* expression (i=10). The ten dots in the figure denote $\mathbf{d}_j^{c=Surprise,i}$ for *Surprise* expression with different intensity $i \in \{1, 2, ..., 9, 10\}$.

For the sake of clarity of presentation, we omit the index j of j_{th} marker unless it is necessary to show it explicitly. In order to learn the category-dependent and intensity-dependent magnification factor $\beta^{c,i}\Delta\mathbf{d}^{c,i}$, we firstly need to obtain exaggerated facial expressions template $\mathbf{d}^{c,APEX}$ with known category c and manually defined apex deformation and subtle expression template $\mathbf{d}^{c,i}$ with known category c and manually defined intensity i from the training database. Based on the prepared facial expression $\mathbf{d}^{c,APEX}$ and subtle expression $\mathbf{d}^{c,i}$ and Eq. 1, we can learn various category-dependent and intensity-dependent magnification factors $\beta^{c,i}\Delta\mathbf{d}^{c,i}$. The mathematical formulation of learning magnification factor $\beta^{c,i}\Delta\mathbf{d}^{c,i}$ is defined bellow:

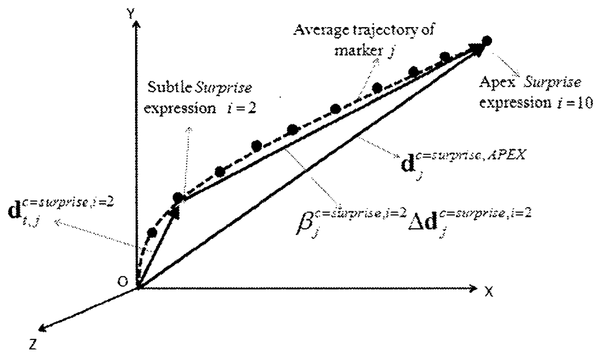

Figure 3. Illustration of learning magnification factor $\beta_j^{c,i}\Delta\mathbf{d}_j^{c,i}$

$$\beta^{c,i}\Delta\mathbf{d}^{c,i} = \overline{\mathbf{d}^{c,APEX}} - \overline{\mathbf{d}^{c,i}} \tag{2}$$

Next, we firstly present how to obtain facial expression template $\overline{\mathbf{d}^{c,APEX}}$ and subtle facial expression template $\overline{\mathbf{d}^{c,i}}$ from training database. And then describe how to learn magnification factor $\beta^{c,i}\Delta\mathbf{d}^{c,i}$.

2.2.1 Calibration of $\mathbf{d}_j^{c,APEX}$ and $\mathbf{d}_j^{c,i}$ from training database

According to Equation 2, we need to obtain exaggerated facial expressions $\mathbf{d}^{c,APEX}$ with known category c and manually defined apex deformation and subtle expression $\mathbf{d}^{c,i}$ with known category c and known intensity i from the training database before learning magnification factor $\beta_j^{c,i}\Delta\mathbf{d}_j^{c,i}$.

For calibration of exaggerated facial expressions $\mathbf{d}^{c,APEX}$, we manually select some exaggerated facial expression candidates with apex deformation for each expression category c and then average them. We consider the average result $\overline{\mathbf{d}^{c,APEX}}$ as the calibration of the exaggerated facial expressions $\mathbf{d}^{c,APEX}$ and use them to learn magnification factor $\beta^{c,i}\Delta\mathbf{d}^{c,i}$.

Based on the motion capture system, we obtain M facial markers' motion trajectories for each sequence from the training database. Similarly to Wang et al. [6], we use B-spline curves to fit the facial markers' motion trajectories. We manually divide each fitting motion trajectory into ten intensity levels ($i \in \{1,2,...,9,10\}$). Then we pick up some facial expression candidates with certain intensity level for each expression category c and average them. We record the average result $\overline{\mathbf{d}^{c,i}}$ for each expression category c and each expression intensity level i, and consider them as the calibration of the facial expressions $\mathbf{d}^{c,i}$. After that, we can use the average result $\overline{\mathbf{d}^{c,i}}$ to learn magnification factor $\beta^{c,i}\Delta\mathbf{d}^{c,i}$.

2.2.2 Learning magnification factor $\beta^{c,i}\Delta\mathbf{d}^{c,i}$

According to Equation 2, we can compute magnification factor $\beta^{c,i}\Delta\mathbf{d}^{c,i}$ defined as below:

$$\beta^{c,i} = \left\| \overline{\mathbf{d}^{c,APEX}} - \overline{\mathbf{d}^{c,i}} \right\| \tag{3}$$

and

$$\Delta\mathbf{d}^{c,i} = (\overline{\mathbf{d}^{c,APEX}} - \overline{\mathbf{d}^{c,i}}) \Big/ \left\| \overline{\mathbf{d}^{c,APEX}} - \overline{\mathbf{d}^{c,i}} \right\| \tag{4}$$

We record the learned category-dependent and intensity-dependent magnification factors $\beta^{c,i}\Delta\mathbf{d}^{c,i}$ corresponds to a certain expression category and intensity hypothesis of pre-magnified expression, and then use then to magnify test subtle expression.

2.3 Test stage

2.3.1 Facial motion magnification

Given a test subtle facial expression $\mathbf{D}_t$, we search all possible magnified expressions by magnifying the test subtle expression $\mathbf{D}_t$ based on all learned magnification factor $\beta^{c,i}\Delta\mathbf{d}^{c,i}$ in Section 2.2.2. The motion magnification of the j_{th} facial marker for test expression $\mathbf{D}_t$ is as defined below:

$$\hat{\mathbf{d}}_{i,j}^{c,APEX} = \mathbf{d}_{t,j} + \beta_j^{c,i}\Delta\mathbf{d}_j^{c,i} \tag{5}$$

Here $\hat{\mathbf{d}}_{i,j}^{c,APEX}$ is the magnified result of the j_{th} facial marker that is magnified by using magnification factor $\beta_j^{c,i}\Delta\mathbf{d}_j^{c,i}$. The magnified expression for test expression $\mathbf{D}_t$ can be represented as $\hat{\mathbf{D}}_i^{c,APEX} = [\hat{\mathbf{d}}_{i,j_1}^{c,APEX}, \hat{\mathbf{d}}_{i,j_2}^{c,APEX}, ..., \hat{\mathbf{d}}_{i,j_{26}}^{c,APEX}]$.

2.3.2 Subtle expression recognition

In order to recognize the test subtle expression, we need to evaluate all obtained magnified expressions by utilizing the similarity between each magnified expression $\hat{\mathbf{D}}_i^{c,APEX}$ and its corresponding learned exaggerated facial expressions template $\overline{\mathbf{D}^{c,APEX}} = [\overline{\mathbf{d}_{j_1}^{c,APEX}}, \overline{\mathbf{d}_{j_2}^{c,APEX}}, ..., \overline{\mathbf{d}_{j_{26}}^{c,APEX}}]$ for each expression category c in Section 2.2.1. Obviously, the similarity is inversely proportional to the distance between them. Therefore, the score of magnified expression $\hat{\mathbf{D}}_i^{c,APEX}$ can be defined as below:

$$Score(\hat{\mathbf{D}}_i^{c,APEX}) = d(\overline{\mathbf{D}^{c,APEX}}, \hat{\mathbf{D}}_i^{c,APEX})^{-1} \tag{6}$$

where $d(\overline{\mathbf{D}^{c,APEX}}, \hat{\mathbf{D}}_i^{c,APEX}) = \left\| \overline{\mathbf{D}^{c,APEX}} - \hat{\mathbf{D}}_i^{c,APEX} \right\|$ is the function to compute the distance between the learned exaggerated facial expressions $\overline{\mathbf{D}^{c,APEX}}$ and magnified displacement $\hat{\mathbf{D}}_i^{c,APEX}$.

Based on Equation.6 we can evaluate the magnified expression $\hat{\mathbf{D}}_i^{c,APEX}$ generated by the learned magnification factor $\beta^{c,i}\Delta\mathbf{d}^{c,i}$. As mentioned in the beginning of Section 2.2, each learned magnification factor $\beta^{c,i}\Delta\mathbf{d}^{c,i}$ is prepared for subtle facial expression with certain expression category c and certain intensity i. Therefore, through selecting the magnification factor which generates the magnified expression with the highest score, we can infer the expression category and intensity of the test expression $\mathbf{D}_t$. The mathematical formulation is defined as bellow:

$$\langle \hat{c}, \hat{i} \rangle = \arg\max_{c,i} Score(\hat{\mathbf{D}}_i^{c,APEX}) \tag{7}$$

As shown in Equation 7, we can infer the category $\hat{c}$ and intensity $\hat{i}$ of test subtle expression from the magnified expressions.

3. Experimental results

We asked two subjects (one male and one female) to make six universal facial expressions (*Happiness, Anger, Surprise, Sadness, Fear, Disgust*) as realistically as possible multiple times in front of the motion capture system mentioned in Sec 2.1. We collected 30 sequences for each subject and each expression category contained 5 sequences. The expression sequence varied from a neutral expression to an apex expression and then back to a neutral expression.

We conducted a person-dependent experiment to test our method by using this database. For each test, one sequence of one expression category was left out and the remaining sequences were used as training data. Figure 4 plots the expression intensity-based recognition result of the male subject using motion magnif-

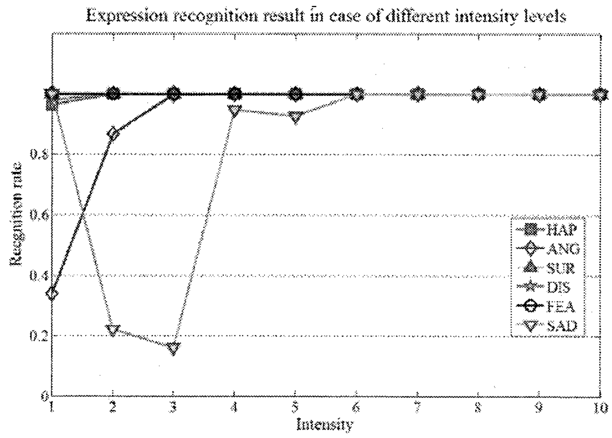

Figure 4. Expression intensity-based recognition result for one subject

ication mentioned in Sec 2.2. As shown in the figure, we obtained good recognition performance with the proposed magnification method. The relatively poor recognition result for *Anger* was caused by the influence of the similar *Fear* expression in the early stage of forming the *Anger* expression. The fact that the *Sadness* expression was poorly recognized at a low intensity level can be attributed to a manual mistake. The subject made two different *Sad* expressions, and their motion feature were different from the common ones in the early stage of the *Sad* expression formation. Therefore, the recognition rate was not better than the other expression categories.

Table 1 compares the average frame-based recognition results of our motion magnification method with a compared method using support vector machines (SVMs) to recognize subtle expressions without motion magnification. For SVMs based method, we firstly manually select some obvious expressions for each expression category from training database and use them to train SVMs for each expression category. The first row of Table 1 shows the recognition result by using the training SVMs to classify subtle expression features without magnification. From Table 1 it is clear that our approach achieves a better recognition performance than the compared method (SVM) without using motion magnification, except for the *Sadness* category. Since the motion features of subtle expressions are intermingled [1], it is difficult to classify them even when using a sophisticated classifier such as SVM. The poorer result of *Sadness* was still due to a manual mistake, as mentioned above.

4. Conclusion and future work

This paper reported a novel approach we developed to recognize subtle expressions using expression category-dependent motion magnification. Motion magnification transforms the subtle facial expressions into corresponding exaggerated expressions by magnifying captured subtle facial deformation by using an expression category-dependent and intensity-dependent magnification factor. The experimental results indicated that the proposed method provided better recognition performance than without using motion magnification and even worked well with exaggerated facial expressions.

In the future, we will try to improve the recognition performance by considering the different contributions of facial markers, using, for example, a weighted facial motion magnification method. The dynamics of expression are also very useful for subtle expression recognition. Last but not least, noise in captured expression data always influences the performance of subtle expression recognition. We will investigate the use of robust feature extraction to reduce the noise.

References

[1] Park, S. and Kim, D: "Subtle facial expression recognition using motion magnification", Pattern Recognition Letters, 30(7), pp. 708- 716, (2009)

[2] Wen, Z. and Huang, T.: "Capturing Subtle Facial Motions in 3D Face Tracking", in Proc. of International Conference on Computer Vision, pp.1343-1350, (2003)

[3] Song, M., Wang, H., Bu, J., Chen, C., and Liu, Z.: "Subtle facial expression modeling with vector field decomposition", in Proc. of International Conference on Image Processing, pp. 2102-2104,(2006)

[4] Cohn, J. F., Zlochower, A. J., Lien, J. J., and Kanade, T: "Feature-Point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expression", in Proc. of the 3rd. International Conference on Face & Gesture Recognition, pp.396- 401, (1998)

[5] Sun, Y. and Yin, L.: "Facial Expression Recognition Based on 3D Dynamic Range Model Sequences", in Proc. of the 10th European Conference on Computer Vision: Part II ", pp. 58-71, (2008).

[6] Wang, M., Iwai, Y., and Yachida, M.: "Expression Recognition from Time-Sequential Facial Images by Use of Expression Change Model", in Proc. of International Conference on Automatic Face and Gesture Recognition, pp. 324-329, (1998).

[7] Liao, W.K. and Medioni, G.: "3D face tracking and expression inference from a 2D sequence using manifold learning", in Proc. of International Conference on Computer Vision and Pattern Recognition, pp. 1-8, (2008).

Table 1 Comparison of frame-based average recognition rates of different methods

	Happiness (%)	Anger (%)	Surprise (%)	Disgust (%)	Fear (%)	Sadness (%)	Average rate (%)
SVM-based method (without magnification)	93.97	80.07	74.40	78.47	76.72	86.32	81.66
Our method (with magnification)	97.35	86.14	98.27	95.39	100	80.59	92.96