

# Enhancing Facial Impression for Video Conference

Sungyeon Park<sup>1,2</sup>, Heeseung Choi<sup>2</sup>, and Ig-Jae Kim<sup>1,2</sup>(✉)

<sup>1</sup> HCI Robotics, University of Science and Technology, Seoul  
Republic of Korea

<sup>2</sup> Imaging Media Research Center, Korea Institute of Science and Technology,  
Seoul, Republic of Korea

{sy.park, hschoi, kij}@imrc.kist.re.kr

**Abstract.** Most people have their preferred impression to be seen by others. Our face warping method can make this real. In this paper, we propose a new technique for automatic transformation of facial impression, for example, look more attractive or baby-face from the original. We build an impression score function, trained by scores from human raters, and the function is used to get displacement vectors for a given face. To preserve the facial identity as much as possible, we constrain the facial deformation range with facial classification. In case of real time application, such as video conference, there are frequent facial expressions variation and position changes. Face tracker is used to cope with this changing situation. Through our experiments, our proposed method can be one of the promising methods for future video conference system.

**Keywords:** Transformation of facial impression · Face warping · Face tracker · Machine learning · Deformation of face

## 1 Introduction

People have developed many kinds of advanced digital technologies that make us enrich all our lives. Among them, smartphone is one of the fascinating inventions. Owing to smartphone, it becomes no surprise that more people can talk to or see others who live in the distant whenever or wherever they are. Since nearly everyone has a smartphone capable of handling video chats, we can enjoy a video chat or teleconference easily. Consequently, we have more chances to talk to someone else face-to-face with digital devices.

When it comes to video conferencing, most people want to look well to others. People have their preferred impression to be seen by others regarding the situation encountered. For instance, we might want to be seen as trustworthy in a case such as job interview and talking with older people. People want to be seen as an attractive person to someone who is in their mind.

The human face conveys important information regarding not only the person's identity but also impressions of various personal attributes from biological ones such as age and gender to social ones such as personality and attractiveness [12]. If the relationship between the physical parameters representing variations in the face's

appearance and the impression of the corresponding images perceived by humans is formalized in a mathematical model, the possibility of a computer capable of dealing with perceptual or subjective information conveyed by faces could become a reality [13].

From this background, we propose a new method to transform facial impression automatically for video chat or teleconference while maintaining identity of original face in this work. Since we transform the impression in regard to the shape of original face, it is possible to maintain close similarity with original face.

In Sect. 2, our transformation facial impression system framework consisting of three stages is presented. In Sects. 3 and 4, we explain details about transformation system and tracking and warping method for video sequence. Finally, we summarize and conclude the chapter in Sects. 5 and 6, respectively.



Fig. 1. System flow chart

## 2 Overview

We showed overview of our proposed method in Fig. 1. Once we capture a person in the frontal view through a camera, we detect the face region and extract 66 landmark points with facial features, such as two eyebrows, two eyes, nose, lips, and boundary of face, from it automatically. In the first stage, we use the detected frontal face image and the landmark points as an input image to the second stage.

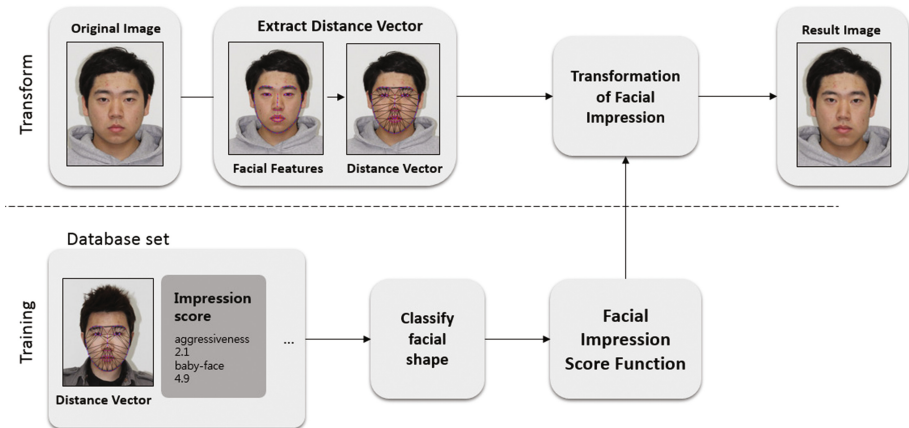


Fig. 2. Detail process for automatic transformation of facial impression

The key process in our system is an automatic transformation of facial impression (Fig. 2) based on our database which we trained. It is composed of frontal faces and their facial impression score that we gathered from groups of human raters. We built *facial impression score function*, modeled by a support vector regressor, with the classified data according to their facial shape. Finally, a deformed face is obtained by optimization of the function. We can transform facial impression as the user wants in this step, in the meanwhile our proposed algorithm can preserve the original face identity.

Facial features are detected in images and tracked across video sequences in the last step. According to the facial features detected through face tracker, original faces are substituted with the deformed face in real time. Our results indicate that the proposed method is capable of transforming facial impression and substituting the original face in real time.

### 3 Impression Transform Engine

#### 3.1 Construction Database and Extract Facial Feature

The training set is consist of 846 frontal portraits of males with neutral expression. The impression (e.g. Attractiveness, baby-face and aggressiveness) degree of each face was rated by 10 human raters. The average rating of a face is its impression score (on a scale of 1 to 7). Since each rater has their own score range and deviation, we use the z-score (Eq. 1) for score normalization which is calculated using the arithmetic mean and standard deviation of the given data [1].

$$s'_k = \frac{s_k - \mu}{\sigma} \quad (1)$$

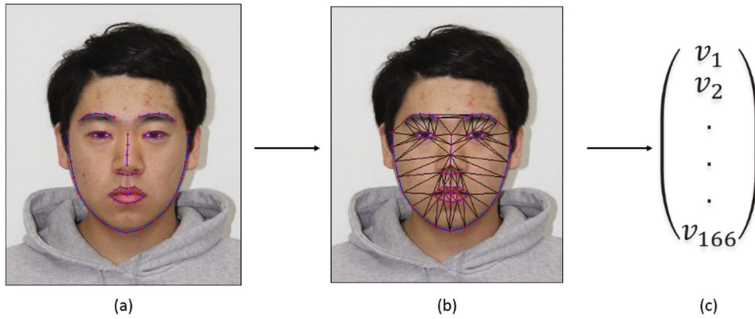
In our work, we use constrained local model (CLM) [11] and Chehra [10] face tracker to extract 66-landmark points (Fig. 3(a)). Since Chehra face tracker which is a good face alignment technique does not provide feature points for the boundary of a face, CLM is used to extract 17-landmark points for the boundary. The extracted feature points are used to construct a Delaunay triangulation which builds meshes (Fig. 3(b)). The triangulation consists of 166 edges, and the lengths of these edges are components of 166-dimensional distance vector (Fig. 3(c)). The distance vector is normalized by the square root of the face area.

**Table 1.** The number of data in each class

	Class 1	Class 2	Class 3	Class 4	Class 5
# of data	116	173	177	187	193

Face area is a sum of the all triangles from meshes. Training samples are classified as their facial shape in five classes by K-Means [8] classifier. Table 1 is the number of data in each class. Classified samples reduce computation and are used to build

Impression Score Function (Sect. 3.2). Constructing function through data which has similar facial shape with original face makes the result images preserve the original facial identity.



**Fig. 3.** (a) Facial feature points, (b) The meshes by Delaunay Triangulation, (c) The 166-D distance vector

### 3.2 Facial Impression Score Function

The image captured by user is original image as input. To seek suitable training sample class for original image, we should select training sample class which has the smallest Euclidean Distances between the center vector in each class and distance vector of original data. Impression Score Estimation function is built from training sample data consisted of selected distance vector and impression score by using Support Vector Regression (SVR) [4].

SVR is similar to Support Vector Machine (SVM) but used for data regression. SVR is an induction algorithm for fitting multidimensional data by using various kernels. Suppose we have training data  $\{(x_1, y_1), \dots, (x_l, y_l)\}$ , where  $x \in \mathbb{R}^d$  and  $y \in \mathbb{R}$ . In  $\epsilon$ -SV regression [1], the goal is to find a function  $f(x)$  that has at most  $\epsilon$  deviation from the actually obtained targets  $y_i$  for all the training data, and at the same time is as flat as possible. Describe the case of linear functions  $f(x)$ , taking the form

$$f(x) = \langle w, u \rangle + b \tag{2}$$

where  $\langle \cdot, \cdot \rangle$  denotes the dot product. Flatness in the case of (Eq. 2) means that one seeks a small  $w$ . For small  $w$ , it is required to minimize the Euclidean norm i.e.  $\| w \|^2$ . This can be written as a convex optimization problem by requiring. We use a Radial Basis Function kernel, which is a popular kernel function used in various kernel method learning algorithms. Model selection was performed by a grid search over the width of the kernel  $\sigma$ , the slack parameter  $C$  and the tube width parameter  $\epsilon$ . Leave-One-Out Cross Validation (LOOCV) method was used to determine adequate parameters. In the (1),  $u$  is 166-dimensional feature vectors and  $y$  is their corresponding impression score. Therefore, we can estimate impression score with above function.

### 3.3 Transformation of Facial Impression

**Reduce Dimension and Constrain the Search Space with PCA.** In this paper, we use Principal Component Analysis (PCA) to reduce a dimension of distance vector from 166 to 30.  $x$  is defined as the distance vector reduced. We regularize objective function (Eq. 3) to constrain the search space in valid human face.  $LP(X)$  (Eq. 4) define face space by multivariate Gaussian distribution. We set  $\alpha$  to 0.3 following experimentation.

$$E(x) = -f(x) - \alpha(f(x) - LP(x)) \quad (3)$$

$$LP(x) = \sum_{i=1}^d \frac{(x - \bar{\mu})^2}{2 \sum_{ij}} + cons \quad (4)$$

In  $LP(X)$ ,  $\bar{\mu}$  is 30-D average vector of training sample data for selected class and  $\sum_{ij}$  is eigen-value obtained by PCA.

**Modified Distance Vector.** We use Powell's method [2] to seek modified distance vector. Powell's method is a proper optimization for problem where the optimal solution is close to the starting point. Since Modified distance vector should be near to original one to maintain close similarity with original, Powell's method is suitable one. Modified distance vector is obtained by minimizing objective function (Eq. 3) which is based on the impression score function. Original distance vector is starting point in optimization.

$$f(x') > f(x) \quad (5)$$

Modified distance vector should satisfy the condition (5). This means that modified data has higher score than original one.

**Adjust Landmark Points as Modified Distance Vector.** To transform the facial impression, landmark points should be placed to new landmark position according to the modified distance vector. New landmark points can be obtained by minimizing Eq. 6. This minimizing reduces the differences between the lengths of these edges and their corresponding modified distances. The Levenberg-Marquardt [5, 6] algorithm is used to perform this minimization.

$$D(q_1, \dots, q_{166}) = \sum \alpha_{ij} (\|q_i - q_j\|^2 - d_{ij}^2)^2 \quad (6)$$

The variable  $q_i$  and  $q_j$  are target landmark point and  $d_{ij}$  is distance corresponded to  $q_i$  and  $q_j$ . We can obtain optimal landmark positions by considering global error.

**Post-processing.** To reduce distortion of deformed face with forehead and background, modified landmark points of face boundary is replaced original one. The modified landmark points  $\{q_i\}$  should translate to the position of the original face. The reason is that it makes deformed face align in the background of the video. Post-processing can cause a decrease in target impression score, but it is necessary to realistic result image.

## 4 Face Tracking and Warping

We should replace original face image to deformed one in real time. For this, we detect face landmark points with Chehra Face Tracker and CLM in real time, and then add the moving amount to deformed landmark points to track the moving face.

We have the set of the original face landmark points  $\{p_i\}$ , which extracted on initial frame, and the set of the deformed one  $\{q_i\}$  in initial image position. The set of difference  $\{diff_i\}$  is acquired by subtracting from the set of original face landmark points  $\{p_i\}$  to the set of moving original face one  $\{p'_i\}$  (Eq. 7). The set of moving deformed face landmark points  $\{q'_i\}$  is updated in every moment by sum of  $\{q_i\}$  and  $\{diff_i\}$  (Eq. 8).

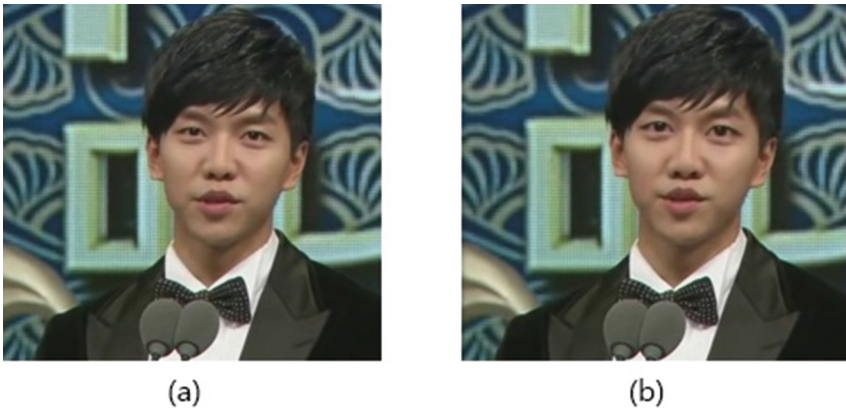
$$diff_i = p_i - p'_i \quad (7)$$

$$q'_i = q_i + diff_i \quad (8)$$

According to the  $\{p'_i\}$  and  $\{q'_i\}$ , we perform texture warping with the triangle mesh which is a result of Delaunay triangulation. The texture images for warping are obtained from each frame to handle the problems of illumination and expression.

## 5 Experiment

We have implemented application for our system. The application automatically detects facial feature in video sequence image. The user looks camera straight with neutral expression and captures frontal image. Next, the application computes for transformation of the facial impression and displays the transformed image. The application track the face and the original face on the video is substituted with deformed faces in real time. Currently, our training sample images are only for Asian



**Fig. 4.** Transformation of facial impression example. (a) original image [14]. (b) deformed image (look much younger)

male. We have deformed face which is not included in training samples to enhance baby-face and transform Fig. 4(a) to (b) based on the training data. The difference between original face and deformed face are subtle on the eyes, nose and lips but this subtle changes have impact on the facial impression (Fig. 5).



**Fig. 5.** The results for transformation of facial impression video sequence images. Top two lines are original images and bottom two lines are deformed images to enhance baby-face. Asian male original and deformed face video sequence images [14].

## 6 Conclusion

In this paper, we propose a new system for automatic transformation of facial impression in real-time. Since we use trained dataset which is classified as their facial shape, the deformed face is close to original face. Since our database is composed of images of Asian male only, so we need to gather additional datasets, such as images for female and other races. In the future, we will gather more review scores for various kinds of impression for more wide use. And besides, we have to come up with a more sophisticated and objective method for evaluating impression than the current one based on human evaluators.

**Acknowledgement.** This work was supported by the KIST Institutional Program (Project No. 2E25660).

## References

1. Jain, A., Nandakumar, K., Ross, A.: Score normalization in multimodal biometric systems. *Pattern Recogn.* **38**, 2270–2285 (2005)
2. Press, W.H., Flannery, B.P., Teukolasky, S.A., Vetterling, W.T.: *Numerical Recipes: The Art of Scientific Computing*, 2nd edn. Cambridge University Press, Cambridge (1992)
3. Eisenthal, Y., Dror, G., Ruppin, E.: Facial attractiveness: beauty and the machine. *Neural Comput.* **18**(1), 119–142 (2006)
4. Joachims, T.: Making Large-scale SVM Learning Practical. In: Schölkopf, B., Burges, C.J.C., Smola, A.J. (eds.) *Advances in Kernel Methods: Support Vector Learning*, pp. 169–184. MIT Press, Cambridge (1999)
5. Levenberg, K.: A method for the solution of certain problems in least squares. *Quart. Appl. Math.* **2**, 164–168 (1944)
6. Marquardt, D.: An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.* **11**(2), 431–441 (1963)
7. Leyvand, T., Cohen-Or, D., Dror, G., Lischinski, D.: Data-driven enhancement of facial attractiveness. *ACM Trans. Graph. SIGGRAPH* **27**(3), 38 (2008)
8. Wagstaff, K., Rogers, S.: Constrained K-means clustering with background knowledge. In: *Proceedings of the 18th International Conference on Machine Learning*, pp. 577–584 (2001)
9. Milborrow, S., Nicolls, F.: Locating facial features with an extended active shape model. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part IV*. LNCS, vol. 5305, pp. 504–513. Springer, Heidelberg (2008)
10. Asthana, A., Zafeiriou, S., Cheng, S., Pantic, M.: Incremental face alignment in the wild. In: *CVPR* (2014)
11. Cristinacce, D., Cootes, T.F.: Feature detection and tracking with constrained local models. In: *Proceeding of British Machine Vision Conference*, vol. 3, pp. 929–938 (2006)
12. Bruce, V.: *Recognizing Faces*. Lawrence Erlbaum Associates, London (1988)
13. Hashimoto, S.: KANSEI as the third target of information processing and related topics in Japan. In: *Proceeding of KANSEI – The Technology of Emotion, AIMI International Workshop*, pp. 101–104, October 1997
14. <https://www.youtube.com/watch?v=sPaRCtGu9zQ>