FACIAL EXPRESSION SYNTHESIS USING GEOMETRICAL WARPING

REPORT

A dissertation submitted in partial fulfillment of the requirements for the degree of Master in the College of Engineering at the University of Kentucky

By
Wanxin Xu
Lexington, Kentucky

Director: Dr. Sen-ching Samson Cheung, Professor of Electrical and Computer Engineering
Lexington, Kentucky

2013
Copyright © Wanxin Xu 2013
FACIAL EXPRESSION SYNTHESIZATION USING GEOMETRICAL WARPING

Facial expression transfer has long been an important topic in computer graphics and vision, driven by applications in character animation, computer games, advertising and more recently, healthcare. However, how to make the synthesized facial expression realistic remains a challenge due to the complexity of human facial anatomy and our inherent sensitivity to facial expression.

The current thesis describes a system that automates facial expression transfer. The process consists of two main parts: feature points detection and geometrical deformation. Feature points detection is achieved by using Active Shape Model (ASM). Inaccurate feature locations in ASM are subsequently corrected by recursively searching for the maximum gradient in local area. This is followed by geometrical deformation in which geometrical warping parameters estimated between feature points from the source faces are applied to the target face. The experimental results show that this method works well on still frontal facial expressions.

KEYWORDS: Active Shape Model, Feature points, Geometrical warping, Facial expression transfer, edge detection
FACIAL EXPRESSION SYNTHESIZATION USING GEOMETRICAL WARping

By

Wanxin Xu

Sen-ching Samson Cheung
Director of Dissertation

Zhi Chen
Director of Graduate Studies

April 2, 2013
ACKNOWLEDGEMENT

This thesis is not possible without the support, guidance from so many people around me.

First of all, I would like to thank my advisor Dr. Sen-ching Samson Cheung. I am deeply impressed for his enthusiasm in research and ability to come up with new ideas. His integrity and motivation also influenced me a lot.

Furthermore, I would like to thank my other thesis committee members, Prof. Daniel Lau and Prof. Qiang Ye, for their time and guidance on my thesis. In addition, I benefited a lot from the courses they taught.

Finally, I would like to thank my parents for their love.
# Table of Contents

ACKNOWLEDGEMENT ........................................................................................................ iii  
Table of Contents ........................................................................................................ iv  
List of Tables ............................................................................................................... v  
List of Figures ............................................................................................................. vi  
Chapter 1 Introduction ................................................................................................... 1  
1.1 Background ........................................................................................................... 1  
1.2 Related work ....................................................................................................... 2  
1.3 Aim of this project .............................................................................................. 3  
1.4 Overview .............................................................................................................. 4  
Chapter 2 Overview of the system ............................................................................. 5  
Chapter 3 Theory ......................................................................................................... 7  
3.1 Active Shape Model .......................................................................................... 7  
  3.1.1 ASM Model Training Stage ...................................................................... 7  
  3.1.2 ASM Searching Stage ............................................................................. 12  
3.2 Facial Expression Transfer ............................................................................... 13  
Chapter 4 Experiment Result and Discussion .......................................................... 17  
4.1 Facial Feature Points Detection .................................................................... 17  
4.2 Facial Expression Transfer .......................................................................... 18  
4.3 Evaluation and improvement ...................................................................... 18  
Chapter 5 Conclusions and Future Work ............................................................... 25  
References ................................................................................................................. 26  
Vita .............................................................................................................................. 27
List of Tables

Table 4.1 The speed for facial expression transfer before and after using MATLAB Coder
............................................................................................................................................. 20
List of Figures

Figure 2.1 Overview of the system. ................................................................. 6
Figure 3.1. Feature points location. Good location (Middle) and Bad location (Right)... 15
Figure 3.2 Good Choice for Landmarks ................................................................ 15
Figure 3.3 Sample image from training set.......................................................... 15
Figure 3.4 The unaligned profiles (Left) and aligned profiles (Right) ..................... 16
Figure 3.5 Facial profile with different $b$ ........................................................... 16
Figure 3.6 Triangulated facial expression.............................................................. 16
Figure 4.1 Frontal face. Input image(Left) and ASM detection result(Right) .......... 21
Figure 4.2 Face with small angle. Input image (Left) and ASM detection result (Right) 21
Figure 4.3 Joker expression. Input image(Left) and ASM detection result(Right) .... 22
Figure 4.4 Triangulated source neutral and target expression .............................. 22
Figure 4.5 Synthesized expression. Input image(Left) and generated expression(Right) 23
Figure 4.6 The average error for ASM. ............................................................... 24
Chapter 1 Introduction

1.1 Background

Although language is the predominant mean by which human beings communicate with each other, facial expressions can sometimes expand our capacity to express meanings that language cannot convey. Therefore, how to synthesize realistic images of human faces has been a fascinating yet challenging problem in computer graphics due to the complexity of human facial anatomy and our inherent sensitivity to facial expression. Analysis and synthesis of facial expressions have been in the spotlight since 1970s and have attracted many researchers from different fields. The pursuit of different solutions for these problems, and their applications in various areas, have led to the advancement of facial expression transfer, that is, to map one person’s expression to another one.

An important step of facial expression transfer is to automatically and accurately detect the feature points, such as eyebrows, eyes, nose, mouth, of the facial expression. This is an easy task for human being, but remains a difficult task for a computer to automatically and quickly specify the required feature points of the face. Researchers in computer graphics and vision have proposed various algorithms to detect feature points of face. These algorithms can be categorized based on the information used: geometry-based, appearance-based, knowledge-based and 3D vision-based [1]. Each category comprises several subcategories. The Active Shape Model (ASM) is such a geometry-based approach that uses a statistical representation of the shapes that are iteratively fitted to a new image.
Once the feature points of face are located, the next step is to find facial changes and apply it to another person to transfer his or her facial expression. Usually, this process can be done by skilled artists manually using kinds of software and the quality of synthesized facial expression is reasonably good. But in general, it is time-consuming and not scalable to high-throughput data like videos. Therefore, a number of investigations have been made to search for methodology to automatically transfer the face expression from one person to another. Such technology can be used in film industry, for example, animated character in Avatar, also, it is of interest in healthcare for the purpose of helping people with disability of emotion recognition. The ability to recognize and understand the emotional meaning from the expressions is necessary step for people to communicate with each other. Studies in [2] show that people with disability of emotion recognition performs well to recognize the expression in the images of themselves. Therefore, the work to transfer expression to the face of people with disability of emotion recognition has potential to improve the outcome of medical care for those patients.

1.2 Related work

As stated above, face feature points detection is an essential step for facial expression transfer, with the purpose of locating and extracting face areas, and useful to capture the saliency of facial expression from the still image or video sequence. Geometry-based methods assume some invariance in the spatial correlations of the face parts. These correlations can be distances, angles, etc [1]. Active Shape Model (ASM) was proposed by Coots et al [3], and can be used for image segmentation, object recognition and image reconstruction. It performs well in shape localization. The detail about ASM will be discussed in the next section. Active Appearance Model (AAM) [4] can be viewed as an
extension of ASM which combines constraints on both shape and texture to define the face appearance.

The transfer of facial expression from one person to another one has been a fascinating yet challenging problem in computer graphics and visions. Since the pioneering work of Parke in 1972, extensive attention has been attracted from both academia and industry, and significant research efforts have been attempted to generate realistic synthesized facial expression. The method used for addressing this problem can be roughly classified into four categories: Parameterization, physics-based approach, performance-driven based approach and data-driven based approach. Williams [5] proposed a system that uses the difference of extracted facial feature to guide the warping. Liu et al. [6] proposed a technique, called Expression Ratio Image, to map the illumination changes of one person’s facial expression to another person so as to improve the mapping result. Zhang et al. [7] developed a system to use geometry component to generate the synthesized expression. Inspired by the previous work, our system relies on the existing technology to transfer the expression from one person to another. The work in my project for facial expression transfer is related to the previous research on facial expression mapping, which by nature is a performance-driven approach. This method can be used in both 2D and 3D cases.

1.3 Aim of this project

In my project, we’ve been seeking a method which allow user to choose a neutral expression from the database, and automatically transfers the chosen expression to the target expression. In order to make the transferred expression more realistic, the detection of feature points on the facial expression needs to be as accurate as possible.
1.4 Overview

The current thesis is primarily focused on the introduction of ASM used as a pre-processing step to extract facial feature points such as eyes, eyebrows, mouth and etc. Also, the process of geometrical deformation to synthesize the expression is discussed in detail. The rest of this thesis is organized as follows: overview of the system to transfer facial expression will be discussed in Chapter 2. The concepts related to the Active Shape Model and geometrical deformation will be presented in Chapter 3. In Chapter 4, experimental results are presented. Conclusion remarks and future work will be given in Chapter 5.
Chapter 2 Overview of the system

In this project, a system to transfer the face from one person to another is discussed. Our system consists of two main parts, facial feature point detection and geometrical deformation. In facial feature point detection, we used Active Shape Model to detect positions (e.g. eyes, eyebrows, nose and mouth) of facial expression. In geometrical deformation, the synthesized expression is produced by warping the specific face areas based on parts of the feature points detected from the previous stage.

The following flow graph in Figure 2.1 shows the data flow of our system. The part marked in red can be viewed as the general procedure of ASM to automatically detect the feature points for a given input image. There are two main stages included by Active Shape Model: training stage and searching stage. The details about those processes will be discussed in the next chapter. The part marked in gray illustrates the transfer from one person’s expression to another one. Such synthesis process will depend on the detection of feature points. Thus, it is important to detect the feature points as accurate as possible. The difference we obtained from the two source expressions will be applied directly to the input facial expression. Finally, we can generate the synthesized expression by using geometrical warping on the target neutral facial expression.

Copyright © Wanxin Xu 2013
Training Set

Aligned using Procrustes Analysis; Apply PCA to shape

Eye center and mouth center detection; Warp mean shape to the target profile

Update the parameters, move the point to the best matched location

Iterate until convergence

Detected feature points

Find difference between two expressions

Geometrical warping

Input image

Synthesized expression

Figure 2.1 Overview of the system.

Copyright © Wanxin Xu 2013
Chapter 3 Theory

Facial feature point location has a great effect on the next stage of facial expression transfer in this project. As an example, Figure 3.1 shows a comparison of a synthesized expression with and without good feature point location. Active Shape Model is one of the classical algorithms to locate the facial feature points [3]. It uses statistical models of shape and appearance to describe an object in the image or sequence. This section first describes the Active Shape Model, followed by a description of the system of facial expression transfer used in this project.

3.1 Active Shape Model

The ASM is first trained based on a set of manually labeled images. After the training, we can use the obtained shape and appearance model to search for the feature points of a given image. Below we present the details of the two parts.

3.1.1 ASM Model Training Stage

In order to locate interesting facial feature points, we need to build a model to represent and describe the shape and variation of the feature points. A set of annotated images is required to build such a statistical model. The first step is to choose a suitable set of landmarks on images, which give a good description about the shape of the face, as the training set. Suppose the number of facial images contained in the training set is \( N \). We must manually label a specific number, say \( n \), of feature points on those images. For the dataset used in the current work, the value of \( N \) and \( n \) are 240 and 58, respectively. So for a 2D image, we can represent the landmark points as a column vector, that is, \( X_i = (x_{i1}, x_{i2}, \ldots, x_{in}, y_{i1}, y_{i2}, \ldots, y_{in})^T \), where \( X_i \) represents the \( i^{th} \) sample from the training set and \( (X_{ik}, Y_{ik}), k=1,2,\ldots,n \) represents the landmark points coordinate. According to
[3], good choices for landmark are points at clear corners of object edge, such as “T” junctions and easily located biological landmarks, as is shown in Figure 3.2. In general, such points are rare and not enough to give an accurate description of the facial shape. Therefore, we add additional points that are equally spaced between the well-defined landmark points. Figure 3.3 shows an example of images with labeled feature points from the training set. To build a comprehensive facial shape model, it is required for training set to contain images with different variations, for instance, expression, sex and pose.

Now we have obtained a set of marked images and their coordinates. In those selected feature points, each point has different variation. Some of them may have big change compared to others, and some points may be more stable. Thus weight for each point is needed to measure their variation.

Take one sample $X_i$ from the training set as an example.

(1) Compute the distance between two arbitrary points. Denote $R_{kl}$ as the distance between the $k^{th}$ and the $l^{th}$ point. If the total number of feature points is 58, then there should be $58 \times 58$ combinations (including the distance of each point to itself).

(2) Compute the variance of each distance obtained from step (1)

(3) Compute the sum of the Variance obtained from step (2) and compute the inverse, the weight of the point can be obtained as

$$w_j = \left(\sum_{i=1}^{n} V_{R_{ij}}\right)^{-1}$$

(3.1)

Where $n$ is the number of the feature points in the sample.
For well-labeled facial image, it is not reasonable to use those points directly to build the model even if we set weight to each of them, due to the difference of size and direction of the image itself. To address this issue and compare the point in the same location from different shapes, we need to align those set of points into a common coordinate. To achieve this, we choose a shape as a reference and then scale, rotate or translate other shapes to align them with the reference shape. After that the mean shape is obtained and normalized. We treat this mean shape as the reference shape and repeat the previous process until the difference between the neighboring mean shapes is less than a specific threshold. The result of unaligned and aligned profiles for the training set is presented in Figure 3.4.

The next step is to use principal component analysis (PCA) [8] to find the statistical information about the shape variation. PCA is an effective approach to reduce the dimension of the data. It plays an important role in feature detection, machine learning, data analysis, etc. It can be defined mathematically as an orthogonal linear transformation that transforms the data to a new coordinate, such that the greatest variance by any projection of the data maps to the first coordinate, the second greatest variance to the second coordinate and so on. In this way, the dimension of data is reduced and most of information is preserved. To improve the efficiency of the ASM, we need to do PCA in the training stage. The detail is as following:

(1) Compute the mean shape and covariance of the training set

\[
\overline{X} = \frac{1}{N} \sum_{i=1}^{N} X_i
\]  
\[
S = \frac{1}{N} \sum_{i=1}^{N} (X_i - \overline{X})(X_i - \overline{X})^T
\]
Where $N$ is number of images in the training set, $X_i$ is the corresponded $i$th shape vector.

(2) Compute the eigenvector and corresponded eigenvalue

We can get eigenvectors and eigenvalues of the covariance matrix $S$ obtained in step (1). According to each eigenvalue, we sort the corresponded eigenvectors as: $\phi_1, \phi_2, \ldots, \phi_n$. The corresponding eigenvalue are: $\lambda_1, \lambda_2, \ldots, \lambda_n$ where $\lambda_i \geq \lambda_{i+1}$ and $i=1,2,\ldots,n$. In order to get the first $t$ eigenvalues,

$$\sum_{i=1}^{t} \lambda_i / \sum_{j=1}^{n} \lambda_j > \eta, \quad \eta = 98\%$$

needs to be satisfied.

(3) Obtain transformation matrix

The obtained transformation matrix is $\phi = \{\phi_1, \phi_2, \ldots, \phi_t\}$. The bigger the eigenvalue, the more important the shape is. Such $t$ eigenvectors are enough to represent the most shape vector $X$.

According to the above steps, that is, applying PCA to the data, we can approximate any shape $X$ in the training set using:

$$X \approx \overline{X} + \phi b$$

(3.4)

Where $b$ is a $t$ dimensional vector given by

$$b = \phi^T (X - \overline{X})$$

(3.5)

As we can see, we can vary the shape $X$ by change the value of $b$ using Equation (3.4). Also, element $b_i$, $i=1,2,\ldots,t$ in $b$ are independent from each other, they can vary in big range. In this case, the obtained face shape will not be the shape we desired. However, to ensure the obtained shape similar to those in the training set, we
must apply limits of \( \pm 3/\sqrt{l_i} \) to each element in vector \( b \). The result of the shape with
different \( b \) is shown in Figure 3.5.

Next, we need to build the appearance model from the training set. Such model
will tell us the distribution of grey-level, which can be referred in the second stage of
ASM to find the best matching point. First, we determine the normal direction of the
feature points in the shape space from the training set, and sample \( k \) pixels along the
either side of each feature point. Then, the grey-level model can be obtained from those
\( 2k + 1 \) pixels, that is, their Mean and Covariance. Suppose their grey-level forms the
vector:

\[
g_{ij} = (g_{ij1}, g_{ij2}, \ldots, g_{ij(2k+1)})^T
\]

(3.6)

Where \( i \) represents the \( ith \) image in the training set \( i = 0, 1, \ldots, N - 1 \), \( j \) represents the
\( jth \) feature points in the image \( j = 0, 1, \ldots, n - 1 \). Get the derivative of the \( g_{ij} \):

\[
dg_{ij} = (g_{ij2} - g_{ij1}, g_{ij3} - g_{ij2}, \ldots, g_{ij(2k+1)} - g_{ij(2k)})^T
\]

(3.7)

Normalize the \( dg_{ij} \):

\[
G_{ij} = dg_{ij} / \left( \sum_{i=1}^{2k} |dg_{ij}| \right)
\]

(3.8)

The mean \( \overline{G}_j \) and covariance matrix \( C_j \) for \( jth \) feature point are:

\[
\overline{G}_j = \frac{1}{N} \sum_{j=1}^{N} G_{ij}
\]

(3.9)

\[
C_j = \frac{1}{N} \sum_{i=1}^{N} (G_{ij} - \overline{G}_j)(G_{ij} - \overline{G}_j)^T
\]

(3.10)
### 3.1.2 ASM Searching Stage

Using the obtained shape and grey-level model from the first stage of ASM, we can search for the feature points on a new facial image. To achieve this automatically, we need to determine the initial position of the face shape as close as possible comparing with the actual landmark position in order to prevent the algorithm from falling into the local minima and also to improve the speed of the convergence. In our implementation, we try to find the initial position of the face shape by detecting the eyes center and mouth center and then warp the mean shape to the input face, using Viola-Jones algorithm [9]. Since points on the initial shape are not always placed on the correct position, we need to find a strategy to move the weak point to the desired position. In ASM, the Mahalanobis Distance is used to choose the best position. Suppose the intensity of point set we obtained using the method to find the grey-level model be \( \{ g_i \} \), the mean and covariance matrix of \( \{ g_i \} \) are \( \bar{g} \) and \( S \) respectively. Let \( g_s \) be the intensity of a possible landmark point, define:

\[
f(g_s) = (g_s - \bar{g})S^{-1}(g_s - \bar{g})^T
\]

(3.11)

One can find the matching position if \( f(g_s) \) reaches minimum.

First, we sample \( m \) pixels on either side of the current point \( (m > k) \) as we did in the previous step. We then select \( 2k + 1 \) pixels from those \( 2m + 1 \) pixels and compare them with the grey-level model we obtained before. One can find the best match from those \( 2(m-k) + 1 \) possible positions using Equation (3.11) and we obtain the change of the shape position \( dX \). To make \( X \) close to \( X + dX \), we can the following:
Let $ds$, $d\theta$, $dt$ be the change of scale, rotation and translation, then the new parameters are $s(1+ds)$, $\theta + d\theta$, $t + dt$. Equation (3.12) can be used to find $dx$.

$$M(s(1+ds); \theta + d\theta)(\overline{X} + \phi b + dx) + (t + dt) = X + dX$$

(3.12)

Where $M(s(1+ds); \theta + d\theta)$ is the transformation matrix and $dx$ can be expressed as:

$$dx = M((s(1+ds))^{-1}, -(\theta + d\theta))[y] - [\overline{X} + \phi b]$$

(3.13)

$$y = M(s, \theta)[\overline{X} + \phi b] + dX - (t + dt)$$

(3.14)

Then, $db$ can be obtained as:

$$db = \phi' dx$$

(3.15)

The final step is to update $s$, $\theta$, $b$, $t$ by:

$$\begin{align*}
s &\rightarrow s(1+ds) \\
\theta &\rightarrow \theta + d\theta \\
b &\rightarrow b + db \\
t &\rightarrow t + dt
\end{align*}$$

(3.16)

The above process is repeated until the shape change is negligible, and we get the desired face shape.

To further improve the efficiency and robustness of the algorithm, one can use a multi-resolution method, which is out of the scope of this thesis and will not be discussed.

### 3.2 Facial Expression Transfer

In this section, we describe the details about facial expression transfer method. Using the detected feature points, one can conduct geometrical deformation to transfer one person’s expression, such as happiness, sadness, angry and etc, to another person via piecewise affine warping. The key to this approach is that the landmarks (the detected feature points) are used to divide the source neutral expression and source target
expression into an equal number of contiguous regions. Delaunay triangulation is a simple and popular way to achieve this, which connects a set of points by a mesh of triangles. Figure 3.6 gives an example of the triangulated facial expression. Using the triangle carried out from the previous step, points within each triangle will mapped to its corresponding coordinates in the output image. According to the points position of the source neutral and source target image, the displacement between two expressions can be calculated and add to the input image.

\[
\begin{bmatrix}
  x_1 & x_2 & x_3 \\
  y_1 & y_2 & y_3 \\
  1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
  \alpha \\
  \beta \\
  \gamma
\end{bmatrix}
= \begin{bmatrix}
  x \\
  y \\
  1
\end{bmatrix}
\quad T\alpha=x
\] (3.17)

\[
\begin{bmatrix}
  x'_1 & x'_2 & x'_3 \\
  y'_1 & y'_2 & y'_3 \\
  1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
  \alpha \\
  \beta \\
  \gamma
\end{bmatrix}
= \begin{bmatrix}
  x' \\
  y' \\
  1
\end{bmatrix}
\quad S\alpha=x'
\] (3.18)

\[
\hat{x}'_i = S\alpha = ST^{-1}x
\] (3.19)

Where \((x_i, y_i), i = 1, 2, 3\) and \((x'_i, y'_i), i = 1, 2, 3\) are coordinates of the vertex on the triangle corresponding to input image and warped image respectively. \((x, y), (x', y')\) are coordinate of any points within the triangle. \(\alpha, \beta, \gamma\) are constant and \(\alpha + \beta + \gamma = 1\).

Equation (3.17), (3.18) can be combined to get Equation (3.19) to map the coordinates from input to output image.

In many cases, piecewise affine warping produces great results, and its computational and conceptual simplicity are further advantages for beginner to do facial expression synthesis.
Figure 3.1. Feature points location. Good location (Middle) and Bad location (Right)

"T" Connection

Figure 3.2 Good Choice for Landmarks

Figure 3.3 Sample image from training set
Figure 3.4 The unaligned profiles (Left) and aligned profiles (Right)

Figure 3.5 Facial profile with different $b$

Figure 3.6 Triangulated facial expression

Copyright © Wanxin Xu 2013
Chapter 4 Experiment Result and Discussion

In this part, the implementation of face feature points detection and facial expression transfer in MATLAB is presented. Experimental results are given and discussed.

4.1 Facial Feature Points Detection

In our implementation, Active Shape Model is used to detect the feature points. To build the engine of ASM, we need a set of training data. In our experiments, we use the “IMM Face Database” [10] as our training data, developed by the Image Analysis & Computer Graphics Group at the Technical University of Denmark, DTU. There are 240 images of faces contained in this database, and each image is annotated by 58 feature points, including eyes, eyebrows, mouth, nose and chin. The information about x-coordinate, y-coordinate, path, type, point number, connects from, connects to of each point are all stored in a separate ASF file. Using the training data, we can build the model and search for the feature points on a given input image. Some input images used in my project belong to FEI Face Database [11]. Figure 4.1 shows the result of feature point detection for the frontal faces using ASM.

Apart from frontal faces, ASM also performs well for some faces with small angle ($\leq \pm 15^\circ$), the result is shown in Figure 4.2. For some faces with uncommon expressions, the ASM also produces good result, which is shown in Figure 4.3. The overall evaluation of the performance of ASM will be discussed in the third part of this section. The experimental result for facial expression transfer will be presented in the next section.
4.2 Facial Expression Transfer

Using the detected feature points, we can transfer one person’s expression, for example, happiness, to another person. Considering that the people’s expression is almost determined by facial areas, such as eyes, eyebrows and mouth, we will use the detected feature points in those areas to conduct facial expression transfer. As discussed in chapter 3, the detected feature points will be triangulated by Delaunay triangulation. In the experiment, the results of triangulated facial expression from source neutral expression and source target expression are shown in Figure 4.4. Since the transfer of expression will partly depend on the difference between two source expressions, we manually label the feature points on these two reference expressions to make the result more accurate and realistic. The left-top corner shows the triangulated target expression for reference image, which is the expression we aim to transfer the input image to be. The synthesized result for frontal faces and faces with angle less than $\pm 15^\circ$ are shown in Figure 4.5.

4.3 Evaluation and improvement

To evaluate the accuracy of feature point detection, we need to know the exact landmark position on a face. We chose the images in training data for the purpose of evaluation. We evaluated the performance by using average error as proposed in [12], defined as Equation (4.1):

$$
E = \frac{1}{N_0} \sum_{i=1}^{N_0} \left( \frac{1}{n} \sum_{j=1}^{n} dis(P_{ij}, P_{ij}') \right)
$$

Where $N_0$ is the number of images, $n$ is the total number of landmark points (for our experiments, $N_0 = 38, n = 58$). $dis$ is the distance. $P_{ij}$ and $P_{ij}'$ are corresponded exact location and detected location. We chose three kinds of expression from the training data.
set, including smile expression, neutral expression and joker expression, to evaluate the performance of the ASM algorithm in my project. The result of average error is plotted in Figure 4.6. As we can see, our method not performs that well for joker expressions, such as expressions with significant changes in head pose or mouth-open, comparing with the result of neutral and smile expressions.

In addition, the speed of program in facial expression transfer stage is not fast, so we used MATLAB Coder, a package in MATLAB, to improve the speed of it. The improvement of speed can extend the use of this system to video sequence instead of still image. The result of speed comparison is presented in Table 4.1.
Table 4.1 The speed for facial expression transfer before and after using MATLAB Coder

<table>
<thead>
<tr>
<th>Expression</th>
<th>Time Before (second)</th>
<th>Time After (second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>91.054295</td>
<td>4.854970</td>
</tr>
<tr>
<td>2</td>
<td>89.610857</td>
<td>4.613658</td>
</tr>
<tr>
<td>3</td>
<td>99.356239</td>
<td>4.983860</td>
</tr>
<tr>
<td>4</td>
<td>66.195062</td>
<td>3.478521</td>
</tr>
<tr>
<td>5</td>
<td>84.544041</td>
<td>4.345975</td>
</tr>
</tbody>
</table>
Figure 4.1 Frontal face. Input image (Left) and ASM detection result (Right)

Figure 4.2 Face with small angle. Input image (Left) and ASM detection result (Right)
Figure 4.3 Joker expression. Input image (Left) and ASM detection result (Right)

Figure 4.4 Triangulated source neutral and target expression
Figure 4.5 Synthesized expression. Input image (Left) and generated expression (Right)
Figure 4.6 The average error for ASM.
Chapter 5  Conclusions and Future Work

A system of using Active Shape Model (ASM) to detect the feature points of facial expression and then transferring the one person’s expression to another person rely on parts of those detected feature points has been proposed and developed. It has been shown that using ASM for feature point detection works well on frontal facial images and faces with small angle. Also, the method of facial expression transfer based on geometrical warping is computationally efficient and effective in rendering realistic expression. The speed of the whole process is optimized using on MATLAB Coder.

There are still much work to be considered for facial feature point detection and facial expression transfer. For example, how to transfer one person’s expression to others in video instead of still image, and also it would be challenge if people in video have significant changes in head pose.

In addition, geometrical warping fails to consider the effect of illumination and other environmental changes on the facial expression, sometimes the result is not that realistic. We can use Expression Ratio Image to capture the change of illumination to address this problem.

Copyright © Wanxin Xu 2013
References


Vita

Wanxin Xu was born on September 30, 1989 in Beijing, China. She graduated from Beijing 101 high school, Beijing in July of 2007. She received the B.S. degree in Communication Engineering from University of Electronic Science and Technology of China, Cheungdu in July of 2011. Since August of 2011, she has been working as a graduate student at University of Kentucky.