

多表情人脸肖像的自动生成

宋 红¹, 黄小川², 王树良¹

(1. 北京理工大学软件学院, 北京 100081; 2. 北京理工大学计算机学院, 北京 100081)

摘 要: 肖像是一种能够抓住人物特征, 又能隐藏细节, 保留个人隐私的艺术表现形式, 本文提出一种以中性表情人脸作为输入, 自动生成不同表情的肖像算法. 首先利用主动形状模型(Active Shape Model, ASM)提取人脸的关键特征点, 然后根据统计学习获取的不同表情人脸的 FAP(Facial Animation Parameter)规则, 对中性人脸的特征点进行变形, 生成表情三角网格, 将中性人脸图像作为纹理映射到表情人脸网格, 生成表情人脸图像, 最后利用图像的梯度域信息和非真实感渲染技术, 将带表情的人脸生成具有肖像风格的人脸图像. 所生成的多表情人脸肖像效果良好, 能够应用于网络和报刊杂志等非真实感图形学和数字娱乐等领域中.

关键词: 人脸肖像; 多表情; 特征形变

中图分类号: TP391.41

文献标识码: A

文章编号: 0372-2112 (2013) 08-1494-06

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2013.08.007

Automatic Generation of Portraits with Multiple Expressions

SONG Hong¹, HUANG Xiao-chuan², WANG Shu-liang¹

(1. School of Software, Beijing Institute of Technology, Beijing 100081, China; 2. School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China)

Abstract: Portrait is a kind of art form, which can catch the features of a person, hide the details and keep the personal privacy. In this paper, an algorithm for automatic generation of human portrait with multiple expressions is presented. The Active Shape Model is used to extract facial features, and the set of Facial Animation Parameters is used to synthesis the multiple expressions for a neutral face. And with the gradient domain of the image and the Non-Photorealistic Rendering, the portrait-liked face images with multiple expressions can be generated automatically. The generated portraits are effective and vivid. They can be applied in non-photorealistic computer graphics and entertainment, such as video, network, newspapers and magazines.

Key words: portrait; multiple expression; feature deformation

1 引言

随着互联网的高速发展, 尤其是社交网络的出现和智能手机、平板电脑的广泛应用, 与数码照片处理相关的研究也成为关注的热点. 肖像的自动生成技术就是其中的一种典型应用. 传统的肖像绘制需要专业的画家通过现场的临摹, 甚至是后期的艺术加工才能完成, 对于没有通过绘画训练的普通人来说, 会显得比较困难. 而借助于计算机图像与视觉的相关技术, 通过输入一张人脸照片, 同样能得到保留人物特质并具有艺术风格的肖像.

在已有的肖像生成研究中, Sherstinkey^[1]等人采用方向滤波器提取图像中的方向信息, 生成类似于华尔街日报中雕版画风格的肖像. Li^[2]等人提出基于形态处理

的人脸线条肖像的提取. 在此基础上, Chen 等^[3]提出了基于范例的方法, 该方法通过大量数据训练, 在贝叶斯概率模型下, 得到人脸与相应线条画的对应关系, 生成相应的线条肖像. 受制于只有线条, 其生成的肖像风格略显单一. Chiang 等^[4]通过提取人脸特征点, 图像变形, 自动生成模仿某类画家风格的肖像, 但其定义的人脸特征点数量过多, 而且某些特征点需采用手动标记.

在人脸表情合成领域, 近年来, 主要有基于实例的方法^[5]和基于特征的方法^[6]. 基于实例的方法, 主要通过获取的运动向量, 对源人脸变形, 得到相应的表情人脸. 基于特征的方法, 则通过对源模型与目标模型进行分析, 对给定的模型进行相似变换, 生成多表情人脸. 多数方法都需要提前预置好相应的模仿实例, 或者人为规定相应的变换准则, 存在着一定的局限性.

本文提出一种基于自动提取特征点,通过对表情数据库进行建模分析提取表情变换规则,并最终生成表情肖像的方法.该方法可自动生成具有艺术风格的多表情人脸肖像.

2 多表情人脸肖像生成系统架构

本文提出的多表情人脸肖像生成系统框架主要由四个模块构成,如图 1 所示.首先对中性表情人脸图像进行特征点定位,生成对应的人脸三角网格;然后根据统计学习获取的不同表情人脸 FAP 规则,对中性人脸的特征点进行变形,生成表情网格;通过仿射变换,将中性人脸图像作为纹理,对发生变化的区域进行纹理映射,生成表情人脸图像;最后利用梯度域信息,采用相应的滤波器,对图像进行非真实感渲染,生成最终带有表情的人脸肖像.

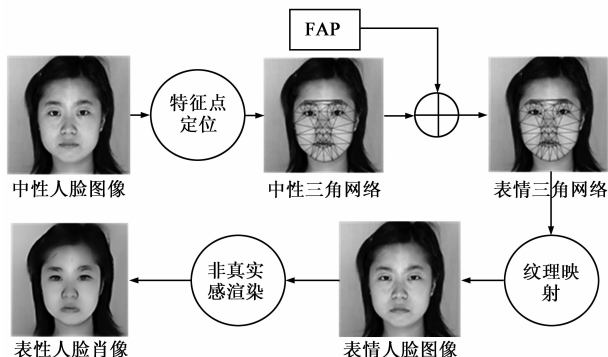


图1 多表情人脸肖像生成系统框架

2.1 基于主动形状模型(ASM)的人脸特征点定位

准确提取人脸特征点是实现表情变换的重要前提和基础.本文采用主动形状模型 ASM^[7,12]进行特征点提取,该模型使用训练样本图像,用一组特征点描述样本的形状信息,然后将样本在同一坐标系下对齐,利用主成分分析(PCA)建立物体的形状模型,最后利用建立的模型,通过一组参数来逼近当前需要匹配对象的形状,从而定位出目标物体.ASM 算法主要包括两个部分:建模过程和匹配过程.

建模过程:

(1)对训练样本中 N 幅图像,每幅图像选取手工标注的 n 个特征点组成特征点集:

$$x_i = (x_i^1, \dots, x_i^n, y_i^1, \dots, y_i^n)^T, i = 1, 2, 3 \dots N \quad (1)$$

(2)对样本集 $x = \{x_i\}$ 进行对齐处理,然后计算训练样本的平均形状向量:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (2)$$

(3)求出协方差矩阵 C ,及其对应的特征值 λ 和特征向量 P

$$C = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T$$

$$\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_{2N}\}$$

$$P = \{P_1, P_2, \dots, P_{2N}\} \quad (3)$$

(4)选取最大的前 t 个特征值 $\lambda' = \{\lambda_1, \lambda_2, \dots, \lambda_t\}$ 和对应的特征值 $P' = \{P_1, P_2, \dots, P_t\}$,则任意的人脸形状

$$X = \bar{x} + \sum_{i=1}^t P_i b_i \quad (4)$$

其中 b_i 是控制特征点形状变化的参数,而 t 一般需要

$$\frac{\sum_{i=1}^t \lambda_i}{\sum_{i=1}^{2N} \lambda_i} \geq 98\%$$

匹配过程:

(1)通过人脸检测方法,提取出人脸区域作为 ASM 输入的初始区域 X .

(2)对每一个特征点建立局部纹理模型,在特征点法线周围的像素亮度的导数进行采样,计算待匹配的特征点与其局部纹理中的像素点欧式距离,最小的即更新为新的特征点位置,最后得到更新的形状 X' .

(3)通过对特征点形状变化参数的控制 ($-\sqrt{3}\lambda_i < b_i < \sqrt{3}\lambda_i$),重新计算新的形状 X'' .

(4)计算 X'' 与上次更新形状 X' 的欧式距离,若小于阈值,则停止匹配,得到最终的形状 X ,否则,转第 2 步继续迭代计算.

本文使用具有 68 个特征点的 ASM 模型,特征点的分布如图 2 所示,能较准确的描述人脸主要器官(眉毛,眼睛,鼻子,嘴巴)的轮廓及特征.

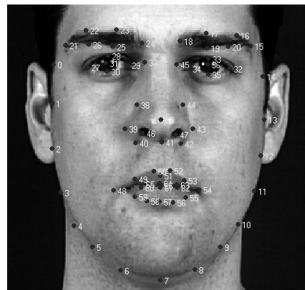


图2 ASM模型68个特征点分布

2.2 人脸表情网格变换

2.2.1 Delaunay 三角网格的生成

为了控制表情生成中特征点变形的幅度,同时考虑到在特征点移动之后能保持原有特征网络的拓扑结构,需要对提取到的人脸特征点,生成相对应的三角网格.本文采用 Delaunay^[8]三角剖分算法生成三角网格,通过 68 个特征点生成三角网格的结果如图 1 中性三角网格所示.

2.2.2 基于 FAP 的表情人脸特征点移动

FAP(Facial Animation Parameter)^[9]是定义在 MPEG-4 标准中的一套脸部动画参数,在三维人脸模型的表情生成以及三维人脸动画上有广泛应用.本文对 FAP 作相应的简化和扩展,使其能应用在二维的中性人脸图像上,生成更自然的表情.在 MPEG-4 定义的脸部动画标准中,主要有 FAP, FAPU, FDP 三个重要参数.

FDP:脸部定义参数(Facial Definition Parameter),主要是定义人脸五官的特征点,通过控制特征点的移动生成动画及表情.在本文中定位的 68 个特征点作为简化的 FDP 特征点,具体特征点分布参照图 2.

FAPU:脸部动画参数单元(Facial Animation Parameter Unit),主要定义脸部动画参数中量度的基本单位,是人脸五官在中性表情下的最基本特征,如嘴巴宽度、眼睛高度等,其分布如图 3 所示.

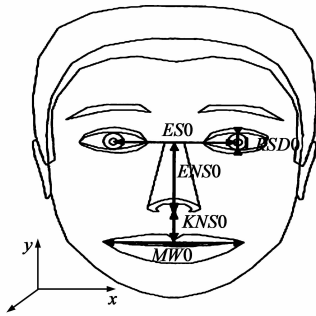


图3 FAPU定义

FAPU 在中性表情上的定义为:

$$\begin{cases} IRISDO = y_{35} - y_{33} \\ MWO = x_{54} - x_{48} \\ MNSO = y_{66} - y_{51} \\ ENSO = y_{67} - y_{37} \\ ESO = x_{36} - x_{31} \end{cases} \quad (5)$$

其中 y_{35} 代表 FDP 中第 35 个特征点在 y 轴上的坐标值. x_{54} 代表 FDP 中第 54 个特征点在 x 轴上的坐标值.

FAP:脸部动画参数,定义了不同脸部特征点进行具体操作的集合,在 MPEG-4 标准的定义中共有 68 组,在文中使用 FAP 来设定特定表情影响特定特征点移动的规则,并针对二维人脸图像的特点对其进行了简化.

为了获取合适的 FAP 参数,文中采用基于统计分析的方法,选取 JAFFE^[10]人脸表情数据库为参照数据库,该数据库中有 10 位女性的 213 张图像,每位女性分别带有生气、厌恶、恐惧、高兴、中性、悲伤、惊讶七种表情.对于数据库中所有中性人脸图像,通过 ASM 方法进行建模,获取平均形状记为 \bar{S}_0 .同样,对于要生成的第 i 种表情,选取数据库中该表情的所有人脸图像,进行 ASM 建模,获取其平均形状记为 \bar{S}_i .

$$\bar{S}_0 = \{P_0^1, P_0^2, \dots, P_0^{68}\}$$

$$\bar{S}_i = \{P_i^1, P_i^2, \dots, P_i^{68}\} \quad (6)$$

其中, P_i^j 表示第 i 种表情的平均形状模型中第 j 个特征点.为了统计 \bar{S}_i 与 \bar{S}_0 之间各对应特征点的移动情况,在同一坐标系下将 \bar{S}_i 和 \bar{S}_0 对齐.在表情变换的过程中,主要关注眉毛、眼睛和嘴巴特征点的移动规律.这些特征点及其对应的 FAP 和 FAPU 具体见表 1 所示.

表 1 FAP, FAPU 分布定义表

FAP	Control FDP	FAPU	Orient
1	48 - 65	MNSO	Y-axis
2	48, 54	MWO	X-axis
3	15 - 26	ENSO	Y-axis
4	18, 24	ESO	X-axis
5	28, 30, 33, 35	IRISDO	Y-axis

将 \bar{S}_0 中第 k 个特征点表示为 P_0^k , 其坐标为 (x_0^k, y_0^k) , 其在 \bar{S}_i 中对应的特征点为 P_i^k , 坐标为 (x_i^k, y_i^k) , 从点 P_0^k 到 P_i^k 的移动可以分解到其坐标在 x 轴和 y 轴上的移动,移动距离用 \bar{S}_0 中计算出的 FAPU 作为量度单位

$$\begin{cases} T_{0,i}^k(x) = \frac{x_i^k - x_0^k}{FAPU_k(x)} \\ T_{0,i}^k(y) = \frac{y_i^k - y_0^k}{FAPU_k(y)} \end{cases} \quad (7)$$

其中 $T_{0,i}^k(x)$ 代表从中性表情到生成第 i 种表情,第 k 个特征点在 x 轴上移动的距离相对于其在这个方向上对应的 FAPU 量度.

对于需要生成多表情的中性人脸图像,通过上面特征点提取步骤得到其形状为 S_0 , 设其第 k 个特征点的坐标为 (x, y) , 其需要生成的第 i 种表情的形状 S_i 中对应的第 k 个特征点的坐标为 (x', y') , 则有

$$\begin{cases} x' = x + T_{0,i}^k(x) \cdot FAPU_k'(x) \cdot \mu_x \\ y' = y + T_{0,i}^k(y) \cdot FAPU_k'(y) \cdot \mu_y \end{cases} \quad (8)$$

其中 $FAPU_k'$ 代表 S_0 中第 k 个特征点的 FAPU 值,引入参数 μ ($\mu \geq 1$) 夸张因子是因为在实际应用中,中性人脸的姿态不一,可能出现某些表情的表达不够明显.通过选取适当的 μ 改变特征点的移动距离,使得生成的表情更为自然明显.

对于表 1 中所有特征点均作上述变换后,对得到的新形状 S_i 进行 Delaunay 三角化,生成对应的表情三角网格.

2.3 人脸表情纹理生成

在生成表情三角网格后,相对应的图像面部纹理像素也要发生相应的变化.本文采用基于三角网格的纹理映射方法.对于表情三角网格中相对于中性三角

网格发生变化的三角面片,通过仿射变换,将中性三角网格中的纹理坐标映射到表情三角网格中,并用其像素值填充。

设表情三角网格中某个三角面片的三个顶点为 $A(x_1, y_1), B(x_2, y_2), C(x_3, y_3)$, 对应中性表情三角网格中的三角面片的三个顶点分别为 $A'(x_1', y_1'), B'(x_2', y_2'), C'(x_3', y_3')$ 。对于 $\triangle ABC$ 来说,其内部的任意一点 $P(x, y)$, 可用 $P = \alpha A + \beta B + \gamma C$ 来描述,其中 (α, β, γ) 称为点 P 相对于 $\triangle ABC$ 的质心坐标,且满足条件: $\alpha + \beta + \gamma = 1$ 。则存在如下关系:

$$\begin{cases} P_x = \alpha x_1 + \beta x_2 + \gamma x_3 \\ P_y = \alpha y_1 + \beta y_2 + \gamma y_3 \\ \alpha = \frac{\text{area}(P, B, C)}{\text{area}(A, B, C)} \\ \beta = \frac{\text{area}(A, P, C)}{\text{area}(A, B, C)} \\ \gamma = \frac{\text{area}(A, B, P)}{\text{area}(A, B, C)} \end{cases} \quad (9)$$

其中 $\text{area}()$ 代表三角形的有向面积,其值定义为

$$\text{area}(T_1, T_2, T_3) = \frac{1}{2} \begin{vmatrix} T_1^x & T_2^x & T_3^x \\ T_1^y & T_2^y & T_3^y \\ 1 & 1 & 1 \end{vmatrix} \quad (10)$$

其方向则按其字母顺序,顺时针旋转为正,逆时针旋转为负。

对于三角仿射变换,根据其不变性要求:点 P 相对于 $\triangle ABC$ 的质心坐标与其仿射变换对应点 P' 相对于 $\triangle A'B'C'$ 的质心坐标相等,则可以得到点 P' 的坐标,如式(11)所示。提取点 P' 在中性人脸图像中的像素值,填充到点 P ,完成对某个点的纹理映射。

$$\begin{cases} P_x' = \alpha x_1' + \beta x_2' + \gamma x_3' \\ P_y' = \alpha y_1' + \beta y_2' + \gamma y_3' \end{cases} \quad (11)$$

对于三角网格中发生变化的三角面片内的每个像素点,作以上仿射变换,填充相应像素值,其余像素直接用中性人脸图像上对应像素填充,即能完成带表情的人脸图像的生成。

2.4 基于梯度域的肖像生成

本文采用二维梯度方法对生成的表情人脸图像作肖像风格化处理。按 Gradientshop^[11] 中的定义,利用二维梯度方法求解问题需要先将问题的输入转化为梯度来表示:

$$F(u, \dots) \rightarrow [d, g, w] \quad (12)$$

其中 F 所带参数表示待处理的图像以及额外的数据, d 为对图像中每个像素值做约束的图像, g 包含 g^x 和 g^y , 分别代表图像每两个像素点之间像素差的参考标准, w 则是 g 和 d 在每一点上权值的集合。根据这些

参数,通过优化能量方程,使得求得的结果图像 f 与 g 和 d 上都能达到接近,具体的能量方程定义为:

$$\begin{cases} E(f) = \sum_{p \in f} E_d(p) + E_g(p) \\ E_d(p) = w^d(p)[f(p) - d(p)]^2 \\ E_g(p) = w^x(p)[f_x(p) - g^x(p)]^2 \\ \quad + w^y(p)[f_y(p) - g^y(p)]^2 \end{cases} \quad (13)$$

在本文肖像风格的生成过程中,对于输入图像,设计如下过滤器:

$$\begin{aligned} F_{stylize}(u, e) &\rightarrow [d, g, w] \\ d(p) &= u(p); w^d(p) = c_1 \\ g^x(p) &= u_x(p) \cdot \cos^2(e^o(p)) \cdot n(p) \\ g^y(p) &= u_y(p) \cdot \sin^2(e^o(p)) \cdot n(p) \\ n(p) &= c_2(1 - e^{-\frac{(e'(p))^2}{2\sigma^2}}) \end{aligned} \quad (14)$$

其中 u 代表输入图像, e 代表图像的长边界检测结果,函数 n 表示通过边缘光亮 $e'(p)$ 得到的抽象夸张度, σ 值越大代表抽象程度越高。参数 c_2 ($c_2 \geq 1$) 控制局部对比度的夸张程度,约束参数 $c_1 \geq 0$ 控制结果图像与源图像的漂移程度。得到输入后,通过上面定义的能量方程,求得最后的解即为具有肖像风格的图像结果。

3 实验结果

本文算法基于 OpenCV 和 C++ 实现,ASM 模型的训练样本为:西安交通大学人工智能与机器人研究所东方人脸库(AI&R)中的 300 张正面人脸图像和 IMM 数据库^[13]中 240 张西方人脸图像,训练获得的模型对人脸特征点提取效果较好。选取训练数据库之外的 4 张中性人脸图像,生成的多表情人脸图像以及多表情肖像结果如图 4 所示。

从实验结果可以看出,本文算法生成的表情人脸图像具有较为明显的表情色彩,最终的结果也具有一定的肖像艺术风格,大体轮廓清晰,局部颜色渐变自然,与源人脸图像真实还原度较高。由于某些特征点在自动提取的过程中不够精确以及每个人的中性表情表达不尽一致,导致在某些局部器官位置(主要是眉毛部分),形变不太精确,结果图像在该部位存在部分失真。

4 总结

本文提出了一种自动生成多表情人脸肖像的方法。采用 ASM 方法提取人脸特征点,通过对表情数据库统计建模,建立一种基于 FAP 机制的表情变换规则对特征点进行表情形变,并通过纹理映射得到多表情人脸图像。在此基础上使用基于梯度域信息的滤波,生成多表情人脸肖像。实验结果表明,该方法在一般的人脸图像上都能取得较为自然、准确的结果。

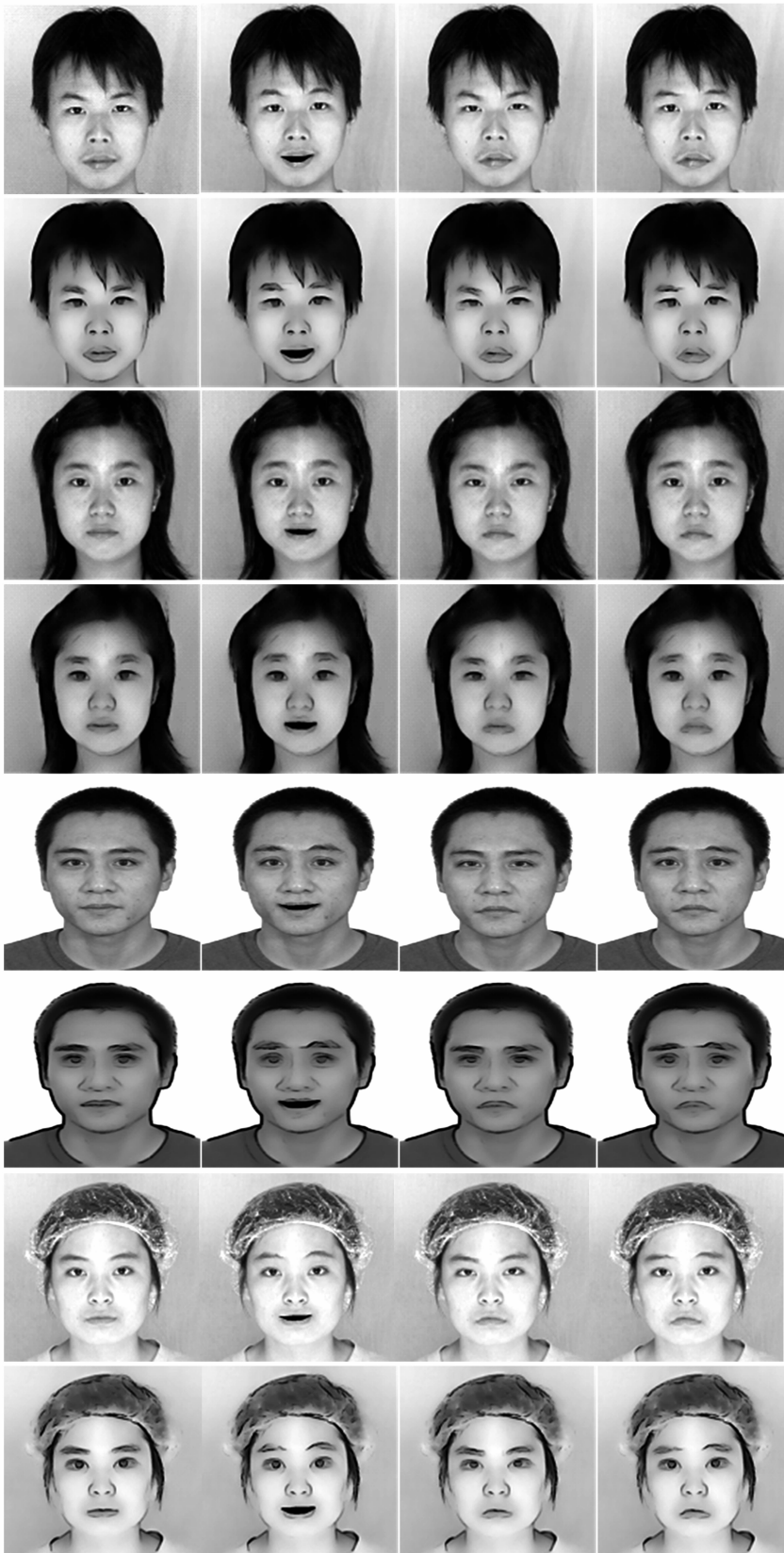


图4 多表情人脸图像及肖像生成

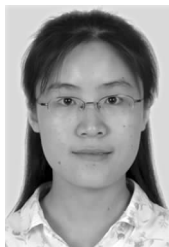
(每组人脸从上到下依次是人脸表情图像、人脸表情肖像, 从左至右, 依次是中性表情、快乐表情、愤怒表情、哀伤表情)

参考文献

- [1] Sherstinsky A, Picard R. M-Lattice: From morphogenesis to image processing [J]. IEEE Transactions on Image Processing. 1996, 5(7): 1137 - 1150.
- [2] Li Y, Kobatake H. Extraction of facial sketch image based on morphological processing [A]. Proceedings of the 1997 IEEE International Conference on Image Processing [C]. Piscataway, NJ: IEEE, 1997. 3: 316 - 319.
- [3] Hong Chen, Ziqiang Liu et al. Example-based composite sketching of human portraits [A]. Proceeding of the 3rd International Symposium on Non-photorealistic Animation and Rendering [C]. New York, NY: ACM, 2004. 95 - 153.
- [4] Pei - Ying Chiang, Wen-Hung Liao, Tsai—Yen Li. Automatic caricature generation by analyzing facial features [A]. In Proceedings of Asian Conference on Computer vision [C]. 2004, Korea.
- [5] P W Hsu, Y Chang, C K. Hsieh, et al. Facial expression cloning: using expressive ratio image and FAP to texture conversion [A]. Proceedings of the 4th International Conference on Information, Communications and Signal Processing [C]. Piscataway, NJ: IEEE, 2003. 2: 694 - 698.
- [6] B Park, H Chung, et al. A feature based approach to facial expression cloning [J]. Computer Animation and Virtual Worlds, 2005, 16: 291 - 303.
- [7] T F Cootes, C J Taylor, et al. Active shape models their training and applications [J]. Computer Vision and Image Understanding, 1995, 61(1): 38 - 59.
- [8] Delaunay Triangulation. [OL] http://en.wikipedia.org/wiki/Delaunay_triangulation, 2012-10-25.
- [9] Tekalp MA, Ostermann J. Face and 2-D mesh animation in MPEG - 4 [J]. Signal Processing: Image Communication, 2000, 15: 387 - 342.

- [10] Matthew N Dailey, Carrie Joyce, et al. Evidence and a computational explanation of cultural differences in facial expression recognition [J]. Emotion, 2010, 10(6): 874 - 893.
- [11] Pravin Bhat C. Lawrence Zitnick, et al. Gradientshop: A gradient-domain optimization framework for image and video filtering [J]. ACM Transactions on Graphics, 2010, 29(2): 10.1 - 10.4.
- [12] 王磊, 邹北骥, 等. 一种改进的提取人脸面部特征点的 AAM 拟合算法 [J]. 电子学报, 2006, 34(8): 1424 - 1427. Wang Lei, Zou Bei, et al. An improved AAM fitting algorithm for extracting human facial features [J]. Acta Electronica Sinica, 2006, 34(8): 1424 - 1427. (in Chinese)
- [13] M. B. Stegmann, B. K. Ersboll, R. Larsen. FAME-a flexible appearance modeling environment [J]. IEEE Trans on Medical Imaging, 2003, 22(10): 1319 - 1331.

作者简介



宋 红 女, 1977 年生于陕西西安. 博士, 副教授. 研究方向为计算机图形图像处理与模式识别.

E-mail: anniesun@bit.edu.cn



黄小川 男, 1988 年生于湖南常德. 硕士研究生, 研究方向为计算机图形图像处理.