

# 基于单张人脸图像的表情生成方法

马禹骞 钟宝江

(苏州大学计算机科学与技术学院, 苏州, 215006)

**摘要:** 人脸表情生成是人机情感交互的重要实现方式之一。通过研究人脸及表情特点, 抽象具体的器官变化, 归纳出针对人脸图像的网格划分法则。在此基础上提出了一种针对人脸表情生成的网格划分技术。该技术通过简单的定位, 使用二次多项式曲线合理划分了人脸主要器官, 继而将其参数化, 实现对面脸各器官的独立调整变动。简化定义了6种基本表情的面部变化, 使用新网格划分技术实现不同的人脸器官变化, 生成各种特定的表情。同时尝试了皱纹等皮肤纹理的添加步骤。与已有的人脸表情生成方法相比, 新方法具有网格划分方式简单、变动参数多样、易于编码实现和计算量小等优点, 最终生成的人脸表情在主观评测中评价良好。

**关键词:** 情感计算; 人脸变形; 表情生成; 网格

**中图分类号:** TP301      **文献标志码:** A

## Expression Synthesis Method Based on Single Facial Image

Ma Yuqian, Zhong Baojiang

(School of Computer Science and Technology, Soochow University, Suzhou, 215006, China)

**Abstract:** Facial expression synthesis is important for affective computing. Here, we investigate various facial and expressions features. By abstracting specific organ changes, several principles for the mesh partition of a human face are provided. A new mesh partition technique is then introduced for the synthesis of human expression. By simply positioning, the new technique divides organs of the human faces with polynomial curves. In consequence, every facial organ can be adjusted separately. The existing definition of six basic facial expressions are simplified. With this mesh generation technique, a new method to generate facial expression is obtained. At the same time, we try a step to add skin texture so that the generated expression can be enhanced. Compared with the existing facial expression synthesis methods, the proposed method is simpler in mesh partition, more flexible in parameter setting, and easier to be coded with less computational cost. Subjective evaluation shows that the generated facial expressions have a good visual effect.

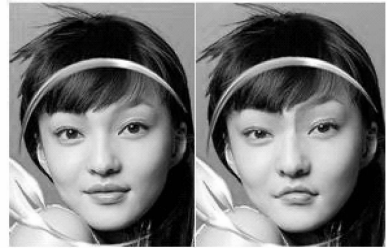
**Key words:** affective computing; face morphing; expression synthesis; meshes

## 引 言

人脸表情生成是人工智能领域的一项有趣挑战, 要求计算机能理解、分析和模拟人类情感, 最终与

人类进行情感交互。相关研究工作的一般关注点主要在人脸识别上,对人脸部分特征的计算机识别率甚至已经高过人类本身。而随着计算机及相关学科不断的发展,人们对人工智能提出了更高的要求。其中,情感计算是近年来越来越热门的研究方向<sup>[1]</sup>,它要求计算机不仅能识别人类情感,还能理解并作出反应,从而像人与人一样进行情感交流。以现有的技术来说,人与机器的交流必须通过屏幕等设备进行,但是机器没有人类多样的表达器官。现在大多数的交互还是以机器为中心的僵硬方式。因而,模拟仿生代理就成实现情感交互的重要媒介。

研究表情及其生成方法,是实现仿生代理、进行情感交互不可或缺的工作。表情生成问题可分为两个部分:人脸建模与表情生成。传统的人脸建模可分为 2D 网格与 3D 网格两种方式。2D 网格建模方法将人脸表情生成问题化归为对图像的变形操作,其基本思想与有限元网格划分类似。比如 Matthews 等<sup>[2]</sup>使用复杂网格,并运用基于统计的方法实现 2D 人脸建模。MPEG-4 标准中也提出了一个不同人脸的动画框架,需要 84 个点划分,这方面 Raouzaoui 等<sup>[3]</sup>做了很多工作。3D 建模需要额外的设备支持,如 Zhang 等<sup>[4]</sup>要求应用一个三维扫描仪,而 Kahler 等<sup>[5]</sup>的方法需要进行大量手工编辑。由于人脸并不规则,通用的建模方式要么划分十分复杂,要么数据获取不便,其余表情生成方法也有相似的困扰。比如 Lee 等<sup>[6]</sup>从生物力学出发,把人脸模型分成 5 层进行变换,使用者需要相当的经验来操纵以获得目标表情;Sifakis 等<sup>[7]</sup>从解剖角度进行表情变换,实际模拟使用了 370 000 个变换单位;Vlasic<sup>[8]</sup>与 Chai 等<sup>[9]</sup>分别从视频角度出发,需要一个包含人脸的视频源输入进行变换;Zhang 等<sup>[4]</sup>使用一组表情源图片,建模后采用线性插值生成表情;Zhang 等<sup>[10]</sup>则直接在图像层面进行插值。这些表情生成方法均需要大量的计算资源或数据,而且常常只能从已有表情出发生成相应表情。近年来,国内学者也广泛开展了关于人脸表情生成及相关研究<sup>[11-14]</sup>。图 1 示例了一种比较成功的表情生成结果。



(a) 原图 (b) 结果  
(a) Original image (b) Result

图 1 一种表情生成方法的结果示例  
Fig. 1 A sample result of one expression synthesis method

现有的表情生成技术,要么数据获取方式十分复杂,要么生成过程复杂,需要大量计算。同时最终效果直接受制于输入表情的丰富程度,总体效果不够自然。特别是在单张人脸图像生成表情方面,相关研究还非常欠缺。本文提出了一种基于单张人脸图像生成表情的方法,该方法针对二维人脸图像,依照人脸及表情变换特征,使用一种新的网格划分方式进行建模。在只有特定对象单张图像的情况下,新方法以较为简单的方式和较小的计算量,取得了良好的表情仿真效果。

## 1 局部表情变化规律

### 1.1 局部器官变化

在局部人脸器官变化的研究方面,Ekman 和 Friesen<sup>[15]</sup>于 1978 年提出了人脸运动编码系统(Facial action coding system, FACS)。该系统基于人脸解剖学理论,依照脸部变化对相关肌肉分组,其中与表情相关的肌肉有 46 组。这项工作本身并不对表情做出识别,而是定义了各种面部动作,因此成为后继人脸表情生成方面研究工作的基础。然而,FACS 系统依照的是人脸实际变动方式,描述过程较为复杂。对于单张图片的表情变换,直接使用该系统会显著增加计算量。为此,本文首先借鉴 FACS 系统的基本思想,依照表情特征来简化变换的分组。

与人脸皮肤光泽等纹理变化相比,人眼对形状变化更为敏感。某些表情变化时,虽然调动的肌肉较

多,但反应在人脸图像上,特别是表情图像的变化上并不明显。在人脸各部分器官中,眉毛、嘴和眼睛变动相对更为显著,因此表情变换单位主要依据这些器官的变换来量化。为了突出形状变化规律,一方面,需要用一定的方式描述其形状轮廓;另一方面,需要描述其形状变化过程。

以下使用日本女性人脸表情数据库(Japanese female facial expressions, JAFFE)作为样本数据。眉毛、嘴和眼睛都可以用贝塞尔曲线勾勒轮廓。为了简化工作,使用较简单的二次曲线进行轮廓勾勒。

如图2所示,在JAFFE数据库中随机选取一些人脸图像,截取出的各类器官都能使用二次曲线进行拟合,这说明使用二次曲线来描述轮廓是合理可行的。在形状变化规律上,眉毛一般以眉心为中心线轴对称呈现为不同程度的拱桥状曲线。不过在表情变化时也会呈现出不对称形状:直线水平或略接近“八”字的形状;眉心处向上抬起,弯曲,呈现出椭圆的部分形状;直线倒“八”字形状。嘴的变动主要呈现为角度与厚度的变化,如厌恶时嘴角向一侧倾斜,悲伤时嘴唇厚度变小。眼睛的变化则主要为眼皮的开合,高兴、悲伤时略微合起;惊讶时张大。



图2 拟合示例

Fig. 2 Fitting examples

## 1.2 人脸网格划分原则

由于本文研究针对单张图像的人脸表情,为了尽量减少不必要的工作,需要依照人脸及表情的变化特点来进行更明确的人脸网络划分。划分的主要原则如下:

### (1) 对称性

一般情况下人脸可以看作一个对称的曲面,因此只需针对一边人脸对象的网格划分,然后对称地运用到另一边。

### (2) 网格密度与变化程度的正相关性

根据一般网格划分法则,需要对曲率大的部分进行细致划分。但是在观察二维人脸图像时,主要对其形状变化差异敏感。因此针对人脸进行网格划分的依据便是各器官部位可能的变动程度:变动程度较大的部分进行细划分,反之进行粗略划分。

### (3) 依据五官特性设计单元网格

人脸各个器官有其各自的特点,应该按照其本身的形状规律和变化可能,进行不同单元的网格形状构造。

### (4) 基于知识进行网格划分

针对个体来说,部分人脸表情的变化是特有的,为此应该进行特殊的网格划分,获得足够恰当的知识经验来划分网格以获得特别的变化。

对人脸进行网格划分可以在每个部分上进行简单的仿射变换或投影变换,反映一类相似或相关的位移、变换,使网格变换更加自然、平滑。为此,从便于编码和计算的角度以比较简单的方式,将人脸从物理世界映射到笛卡尔坐标系,实施后继的操作。

## 2 针对表情的人脸网格划分

文献中已有的划分方式主要从不同角度对人脸做出了细致的描述,需要相当数量的参照或定位点。对于人脸图像来说,没有将主要器官变化与周围皮肤变化区分开来,从而难以实现针对单张人脸图像的多表情生成。为此,本节提出一种新的网格划分方式,即把器官与周围皮肤区分开来,分别对待。如图3所

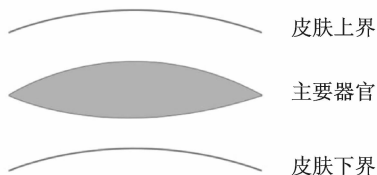


图3 本文所使用的网格划分法示意图

Fig. 3 Demonstration of the mesh partition technique used in the paper

示,在处理人脸图像时,器官部分的图像应该能做更细致、灵活的变化,周围皮肤的变化则简单处理以节省计算量。

根据图 2 所示的依据,可以使用简单的曲线(如二次函数)来描述器官轮廓。眉毛、嘴和眼睛这些关键器官,都可以看作是一个纺锤体或纺锤体的变体,以两条曲线划分出来。为了统一,需要变换的皮肤区域也用二次曲线划分网格。

本文定义,“上皮肤”指皮肤上界与主要器官上界中间的条状区域;“变换对象”指主要器官上界与主要器官下界形成的纺锤形区域;“下皮肤”指主要器官下界,与皮肤下界中间的条状区域。这样,可以将区域划分为上皮肤、变换对象和下皮肤 3 部分。依照以上划分的曲线,把区域再划分成一个个四边形。虽然四边形由直线构成,但是适当的划分数量就能够满足近似曲线的效果。以上的划分将人脸各器官区域划分为上皮肤、变换源区域(变换目标区域)和下皮肤 3 部分。具体地说,每条曲线在横向上做均等分隔。如图 4 所示,从左往右开始,皮肤边界曲线上的点与对象区域边界曲线的上点一一对应,形成多个四边形分隔区域。特别地,图 4 中点 1,4,8;3,7,10;1,11,8;3,13,10 形成三角形。为了区域的统一,可以取 4,7,11,13 这 4 个点所在  $y$  轴方向上的一小段直线,使这 4 个区域的划分仍然为四边形。

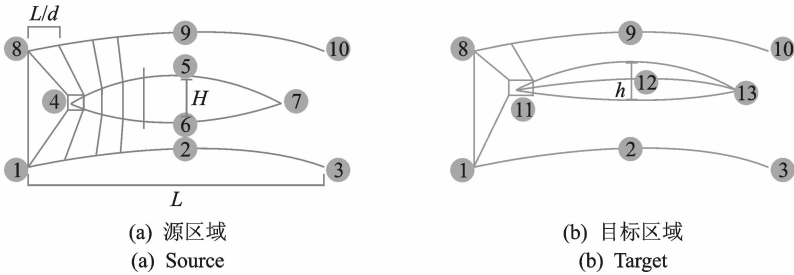


图 4 针对表情的人脸网格划分

Fig. 4 Facial mesh partition for expression

同时,定义  $d$  为划分密度; $r$  为比例系数。由此得到“上皮肤”区域、“变换对象”区域和“下皮肤”区域各  $d$  小块区域,加上左右两个区域,一共需要变换  $3d + 2$  个四边形小块区域。记  $H$  为主要部分原始高度, $h$  为变换后目标区域高度。如图 4 所示,变换后目标只有方向趋势定义(点 11,12,13),其本身没有高度。变换后的高度取决于原高度与比例系数,即  $h = rH$ 。这样对于每个小分块,已知源图像与目标图像 4 个点,只要对应区域,进行投影变换即可。使用以上变换方式,在一张中性人脸图像嘴部进行变换。为减少计算量,参数  $d$  一般取值较小。具体地,由于划分成四边形的投影变换,划分的数量即  $d$  的设置关系到变换效果的平滑程度。如果划分得太少,其效果就会如图 5 所示,所生成的人脸表情很不自然;而划分得过多会加大计算量。在多次试验后发现一般  $d \geq 5$  的划分可以取得良好的效果。

以下列出各部分人脸器官可能发生的变化。

**眉毛** 眉毛部位可以分成 3 类变换方式:内眉抬起、内眉下垂和眉毛抬起。

**嘴** 嘴部可以主要分成 3 类变换方式:嘴角上扬、嘴角下垂和嘴不对称弯曲。其他表情如抿嘴等可以通过调节嘴的厚度,即关键区域的高度来实现。

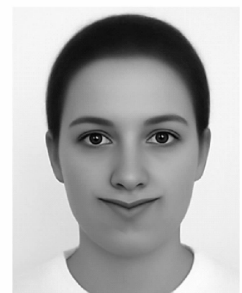


图 5 不合理的参数设置效果  
( $d=2$ )

Fig. 5 Result of unsuitable parameter setting ( $d=2$ )

**眼睛** 眼睛部位的可能变动比较细微,特别是眼球的角度。这里只考虑眼皮的变化,即其高度。

通过上述分析,可以对人脸主要表情相关部分,即眉毛、嘴和眼睛进行独立的调整、控制,见图 6 所示。但表情是人脸的一个综合表现,上述这些变化只是局部的、无感情的器官变化。

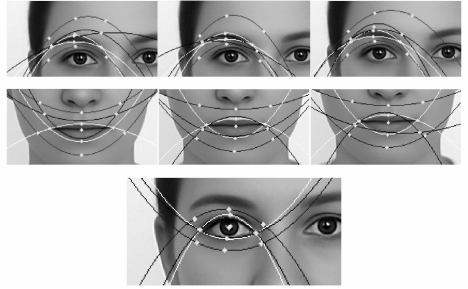


图 6 局部人脸变化示例

Fig. 6 Demonstration of partial facial changes

### 3 多表情生成

#### 3.1 基本表情生成

Ekman 和 Friesen<sup>[16]</sup> 在 1971 年研究了不同种族与文化的人群间表情的情况,提出人类具有 6 种基本表情:高兴、悲伤、愤怒、惊讶、厌恶和恐惧。这些表情在不同种族或文化间都有相似的表达,可以说是人类共有的表情。但是因为表情本身具有主观性,同样的面部表情对不同的人,在不同的情景下可能有不同的语义。所以从表情的基础即脸部变化开始研究,这样只要能表现足够的脸部变化类型,就能生成多种丰富的表情。其中关键的难点在于使用哪些基本模块生成所需要的表情。

依照表情一般规律,以及对 JAFFE 数据库的观察,针对 6 种基本表情,可以做出如表 1 所示的定义。在前面工作的基础上,可以组合这些变动形成一个完整的表情。以“高兴”表情嘴部变化为例(如图 7 所示),白色轮廓为勾勒的源区域,黑色部分为经过参数计算后的目标区域。因为皮肤区域纹理较为一致,主要变换嘴唇区域,在视觉上即能达到明显效果,同时皮肤区域无扭曲感。

表 1 6 种基本表情定义

Tab. 1 Definition of six kinds of basic expressions

#	表情	文字说明
1	高兴	眉毛放松。嘴角往耳朵处拉伸。
2	悲伤	内眉毛弯曲向上。眼睛轻微合上。嘴角下垂。
3	愤怒	内眉毛向下拉伸聚拢。眼睛张大。
4	恐惧	眉毛抬起聚拢。内眉毛弯曲向上。
5	厌恶	眉毛和眼皮放松。上嘴唇抬起、弯曲,通常不对称。
6	惊讶	眉毛抬起。眼皮张开。

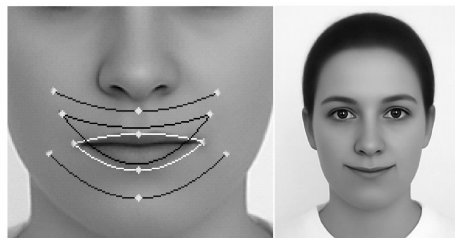


图 7 “高兴”嘴部变化

Fig. 7 Transformation of mouth in happiness

同样地,通过组合每一个器官的变化,就能生成其他表情,其生成效果如图 8 所示。通过不同的组合,也能生成更为复杂细致的表情。

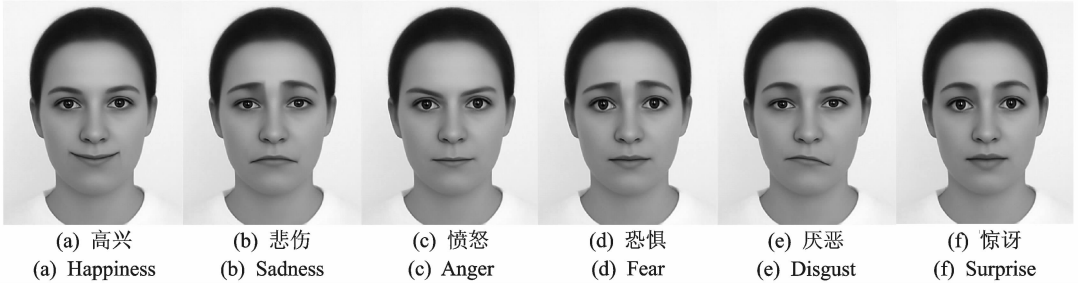


图 8 6 种基本表情的生成

Fig. 8 Generation of six kinds of basic expressions

### 3.2 皮肤纹理的添加

计算机生成的人脸图像之所以常常给人以虚假的感觉,是因为其表面纹理不如真实世界当中的细节丰富。一些变化,如愤怒时眉毛的聚拢,通过简单的移动眉毛,即仿射变换,难以做出真实的效果。但是纹理的自然生成一直是图形图像方面的难点。因为其受光照因素影响很大,需要考虑的因素很多。可以针对单张人脸的情况,提取一些褶皱图像。图 9 为皱眉纹理图像。通过在“愤怒”表情上添加皱眉效果,可以得到更加真实的表情,结果如图 10 所示。

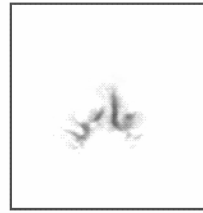


图 9 皱眉纹理

Fig. 9 Wrinkle texture

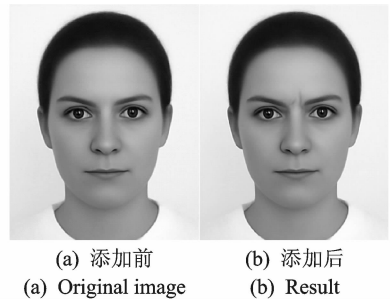
### 4 表情生成的主观评测结果

因为表情没有一个客观的、普适的计算量化标准,难以确定生成的表情图片是否达到了一定水平。为此,针对表情生成结果,进行了人眼主观性测评即问卷调查。问卷分为两个问题,给调查对象 6 幅图片,进行如下测评:(1) 请求其找到对应的表情;(2) 评价表情自然程度,表情自然程度分为 5 级,1 为粗糙,5 为自然。统计结果如表 2 所示。

从表 2 可看出,针对新方法所生成的表情整体识别率都达到了 60% 以上。部分表情,如“高兴”和“愤怒”,识别的正确率达到 100%。在自然度方面的结果与识别正确率不完全对应,如“厌恶”的识别正确率较高,达到了 85.7%,而自然度评分却是几个表情中最低的。另外值得注意的是,添加了皱纹的“愤怒”得到了最高的自然度与识别率,显示了所采用的皮肤褶皱添加方法的必要性。以上测评结果说明新方法具备良好的表情生成效果。

### 5 结束语

人脸多表情生成是人工智能领域一个复杂问题。因为人类视觉对于人脸特别敏感,因此提高生成



(a) 添加前 (b) 添加后  
(a) Original image (b) Result

图 10 添加纹理后愤怒表情比较

Fig. 10 Comparison of the anger expression before and after adding texture

表 2 针对新方法的问卷结果

Tab. 2 Questionnaire result of the proposed method

表情	正确率/%	自然度
高兴	100.0	4.5
悲伤	92.9	3.9
愤怒	100.0	4.8
恐惧	64.3	3.9
厌恶	85.7	2.8
惊讶	71.4	4.0

表情的自然程度非常困难。现有的研究方式多集中在通过昂贵的采集设备,以复杂的计算方式生成人脸多种表情,存在数据获取困难、生成方式复杂、资源消耗量大和生成表情有限等限制。针对人脸具体可调整的参数也十分有限。本文首先提出一种人脸网格划分方式,能够有效量化不同的表情变化,并且讨论了关键变量的一般取值范围,具有数据获取简单,计算量小,生成图片自然等优点。然后研究局部器官变化方法,新方法不受制于已有表情定义,理论上可生成任意表情。在此基础上,通过定义具体表情的变化参数,实现多种表情的生成。同时尝试添加褶皱纹理,最终取得了良好的表情生成效果。主观评测结果显示新方法生成的人脸表情比较自然和真实,具有一定的应用前景。

## 参考文献:

- [1] Picard R W. Affective computing[M]. Cambridge: The MIT Press, 2000.
- [2] Matthews I, Baker S. Active appearance models revisited[J]. International Journal of Computer Vision, 2004, 60(2): 135-164.
- [3] Raouzaoui A, Tsapatsoulis N, Karpouzis K, et al. Parameterized facial expression synthesis based on MPEG-4[J]. EURASIP Journal on Applied Signal Processing, 2002, 2002(1): 1021-1038.
- [4] Zhang L, Snavely N, Curless B, et al. Spacetime faces: High-resolution capture for modeling and animation[C]//Data-Driven 3D Facial Animation. London: Springer, 2008: 248-276.
- [5] Kähler K, Haber J, Seidel H P. Geometry-based muscle modeling for facial animation[C]//Proceedings of the Graphics Interface 2001 Conference. Ottawa, Ontario, Canada: Canadian Human-Computer Communications Society, 2001: 37-46.
- [6] Lee Y, Terzopoulos D, Waters K. Realistic modeling for facial animation[C]//Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM, 1995: 55-62.
- [7] Sifakis E, Neverov I, Fedkiw R. Automatic determination of facial muscle activations from sparse motion capture marker data[C]//ACM Transactions on Graphics (TOG). New York: ACM, 2005, 24(3): 417-425.
- [8] Vlasic D, Brand M, Pfister H, et al. Face transfer with multilinear models[C]//ACM Transactions on Graphics (TOG). New York: ACM, 2005, 24(3): 426-433.
- [9] Chai J, Xiao J, Hodgins J. Vision-based control of 3d facial animation[C]//Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation. San Diego: Eurographics Association, 2003: 193-206.
- [10] Zhang Q, Liu Z, Quo G, et al. Geometry-driven photorealistic facial expression synthesis[J]. Visualization and Computer Graphics, IEEE Transactions on, 2006, 12(1): 48-60.
- [11] 宋红, 黄小川, 王树良. 多表情人脸肖像的自动生成[J]. 电子学报, 2013, 41(8): 1494-1499.  
Song Hong, Huang Xiaochuan, Wang Shuliang. Automatic generation of portraits with multiple expressions[J]. Chinese Journal of Electronics, 2013, 41(8): 1494-1499.
- [12] 谈国新, 孙传明. 一种真实感三维人脸交互式生成方法[J]. 武汉大学学报:信息科学版, 2014, 39(8): 992-997.  
Tan Guoxin, Sun Chuanming. An interactive approach to generate realistic 3D face[J]. Geomatics and Information Science of Wuhan University, 2014, 39(8): 992-997.
- [13] 蒋斌, 贾克斌. 一种用于表情识别的局部判别分量分析算法[J]. 电子学报, 2014, 42(1): 155-159.  
Jiang Bin, Jia Kebin. A local discriminative component analysis algorithm for facial expression recognition[J]. Chinese Journal of Electronics, 2014, 42(1): 155-159.
- [14] 段红帅, 朱振峰, 赵耀. 一种基于双子空间的人脸美感分析方法[J]. 数据采集与处理, 2012, 27(1):105-110.  
Duan Hongshuai, Zhu Zhenfeng, Zhao Yao. A dual subspace algorithm for facial attractiveness analysis[J]. Journal of Data Acquisition and Processing, 2012, 27(1):105-110.
- [15] Ekman P, Friesen W V. Manual for the facial action coding system[M]. California: Consulting Psychologists Press, 1978.
- [16] Ekman P, Friesen W V. Constants across cultures in the face and emotion[J]. Journal of Personality and Social Psychology, 1971, 17(2): 124.

## 作者简介:



马禹骞(1992-),男,工程师,研究方向:计算机视觉和图像分析。



钟宝江(1972-),男,博士,教授,研究方向:计算机视觉、图像分析与理解, E-mail: bjzhong@suda.edu.cn。

