

# 基于图像的个性化卡通系统\*

陈洪<sup>1</sup>, 郑南宁<sup>1</sup>, 梁林<sup>2</sup>, 李岩<sup>2</sup>, 徐迎庆<sup>2</sup>, 沈向洋<sup>2</sup>

<sup>1</sup>(西安交通大学 人工智能与机器人研究所, 陕西 西安 710049);

<sup>2</sup>(微软亚洲研究院, 北京 100080)

E-mail: {hchen, nnzheng}@aiar.xjtu.edu.cn; {i-liang, yli, yqxu, hshum}@microsoft.com

http://www.aiar.xjtu.edu.cn; http://research.microsoft.com/asia/group/research/visualize.asp

**摘要:** 介绍了一个卡通动画系统. 该系统可使用户很方便地生成个性化的卡通头像, 并通过简单的交互定制各种夸张而生动的表情, 还可以实时生成语音驱动的卡通动画. 系统主要包括 3 部分: 基于图像的卡通自动生成模块、交互的卡通表情编辑模块和语音驱动的卡通动画模块. 卡通自动生成模块采用基于样本学习的方法生成具有特定艺术风格的卡通头像. 这里使用了一种非参数化采样算法来学习原始照片和画像间复杂的统计关系. 在此基础上, 卡通表情编辑器可以通过预先设计的模板生成各种具有生动表情的卡通. 同时, 用户也可以定制自己的表情模板. 卡通动画模块使用了实时的唇形合成算法, 可以生成语音驱动的卡通动画. 该系统可以使普通用户很方便地生成个性化的卡通头像和动画, 在这一点上优于其他动画制作系统.

**关键词:** 非照片真实感绘制; 基于样本的学习; 非参数化采样; 唇形合成动画

中图法分类号: TP391 文献标识码: A

卡通画通过轮廓、线条来描绘物体和场景, 是一种独特的艺术表现手法. 但是, 卡通绘制并不容易, 只有受过专业训练的画家才具有这样的技巧. 近年来, 随着计算机技术的发展, 出现了很多技术, 可以使艺术家使用计算机绘画, 其中包括笔划的绘制技术<sup>[1-3]</sup>、色调控制<sup>[2,3]</sup>和各种数字化输入设备. 结合这些技术, 开发出了多种卡通设计和动画系统<sup>[4,5]</sup>. 虽然这些系统中使用很多模板来简化卡通及其动画的编辑, 但未经训练的使用者仍然很难使用它们来生成具有个性化的卡通——带有个人特征的卡通. 而在电子贺卡、网上聊天等应用领域, 生成个性化的卡通及动画是很有意义的.

为了使普通用户很方便地生成个性化的卡通和动画, 我们设计了本文的系统. 根据输入的照片, 系统可以自动生成个性化的卡通, 在此基础上, 可以很方便地编辑出带有特定表情和夸张效果的卡通画, 并自动地生成语音驱动的动作.

卡通的自动生成分为两部分: 生成不带笔划风格的线条画和笔划绘制. 本文使用基于样本学习的方法来自动生成线条画. 非参数化采样用于学习线条画与原始图像间复杂的统计特性, 再结合模板匹配即可从输入图像自动地生成量化的线条画. 通过骨架笔划的技术<sup>[1]</sup>绘制笔划, 生成卡通画.

一旦生成卡通头像, 通过选用系统预先设计的表情模板, 用户即可改变卡通的表情, 产生各种生动的夸张效果. 本系统还采用了一种实时的唇形合成算法来自动地生成语音驱动的卡通动画. 该算法提取底层的语音特征, 适用于各种语言.

\* 收稿日期: 2002-03-05; 修改日期: 2002-06-18

基金项目: 国家自然科学基金资助项目(60024301)

作者简介: 陈洪(1975 - ), 男, 浙江金华人, 博士, 主要研究领域为模式识别与智能系统, 机器视觉与图像处理; 郑南宁(1952 - ), 男, 江苏南京人, 博士, 教授, 博士生导师, 主要研究领域为模式识别与智能系统, 机器视觉与图像处理; 梁林(1974 - ), 女, 广西玉林人, 助理研究员, 主要研究领域为计算机图形学, 机器视觉, 图像处理; 李岩(1973 - ), 男, 河南南阳人, 助理研究员, 主要研究领域为机器视觉, 计算机图形学; 徐迎庆, 男, 博士, 研究员, 主要研究领域为计算机图形学, 机器学习; 沈向洋, 男, 博士, 高级研究员, 主要研究领域为机器视觉, 计算机图形学.

本文第 1 节介绍本文的相关工作.第 2 节介绍系统的整体框架.在接下来的章节中将详细地解释每一个模块的关键算法,并给出系统生成的结果示例.最后对系统的特点进行总结,并进一步讨论系统将来的扩展.

## 1 相关工作

现有的大多数卡通生成系统<sup>[4,5]</sup>主要是提供一些底层的编辑和控制工具来帮助用户绘制卡通并生成动画.例如,Inkwell<sup>[6]</sup>中介绍了一些很有效的动画技术:使用分层、层次化的模块组合、定义运动函数等.最近的 CharToon 系统<sup>[4]</sup>又提供了一些附加的功能:针对人脸定义特定的骨架,利用骨架动画技术可支持模块和动画的重用.但是,这些系统都是为有经验的画家提供动画制作工具,而本文中的系统是要使普通用户也可以很方便地生成个性化的卡通和动画.

在交互生成人脸漫画方面也有一些尝试.Brennan<sup>[7]</sup>提出了一个交互的漫画生成系统.Murakami 等人开发出了一个基于模板的人脸漫画系统 PICASSO<sup>[8]</sup>和在线系统 Web-PICASSO<sup>[9]</sup>.但是,这些方法都没有使用基于样本学习的方法,而是采用自定义的规则,因此生成的漫画不够生动.与这些方法类似,本文的系统也采用交互方式生成夸张的卡通,但是可以使用预定义的夸张模板生成特定风格的卡通,同时还保持了特定人的特征.

在专业的绘画书籍<sup>[10]</sup>中指出了绘画作品的一些关键因素,如形态、透视、轮廓、对细节刻画程度的控制等.文献[3]将线条绘制与更高层次的绘画控制分离开来.用户可以控制整幅画的色调、绘制细节等较高层的属性以及线条放置的具体位置,而繁琐、单调的线条绘制则由计算机来完成.但是,文献[3]所述系统针对的使用者是有经验的艺术家,对于普通用户,还是很难把握这些控制属性.本文的系统也把卡通画的生成分解成两个部分:人脸线条画的生成和线条绘制.与文献[3]不同的是,通过基于样本学习的方法,线条画的生成是自动的,用户不需要有任何绘画经验.

目前还开发了一些从人脸照片生成画像的技术.比如,Sherstinkey 等人使用 M-网格方法自动生成用于《华尔街日报》的雕版画效果<sup>[11]</sup>,Wong 开发了一个半自动肖像画生成系统<sup>[12]</sup>.但是,这些工作都是设计算法来模拟特定的艺术效果,而我们的方法则是通过一组样本来学习特定的艺术风格.

语音驱动的唇形自动合成是动画生成中的重要研究内容<sup>[13,14]</sup>.本文的系统使用这项技术来减轻用户编辑动画的难度.Phoneme-viseme 映射<sup>[13]</sup>是一种被广泛使用且很有效的唇形合成方法,但是这种方法受到语言种类的限制,而且其计算复杂度较高,无法实现动画的实时生成.本文的算法直接采用底层的语音特征驱动唇形,适用于各类语言,并可实时生成动画.

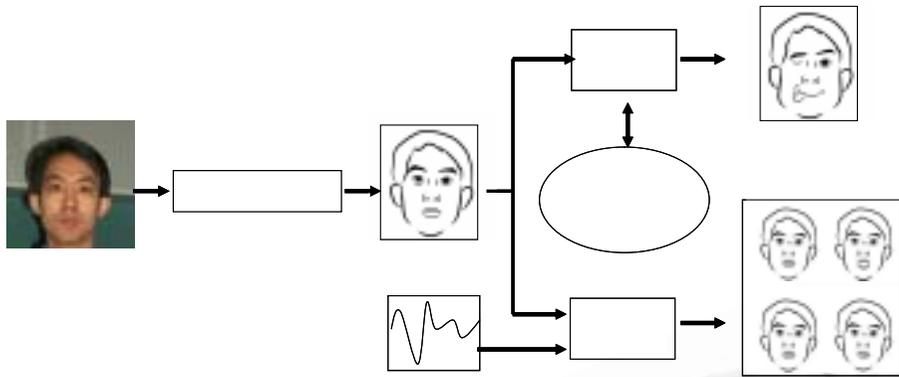
## 2 系统框架

本系统的目的是使普通用户 can 很方便地生成和编辑个性化的卡通头像,并生成语音驱动动画.如图 1 所示,系统主要包括以下 3 个部分:

(1) 卡通生成模块.根据输入的正面人脸照片自动生成卡通头像.该模块生成的卡通未经任何夸张变形,尽量保持原有人脸的特征.系统可生成矢量形式的结果,便于进一步编辑和生成动画.根据用户的选择系统可生成不同线条风格的卡通.卡通生成模块是整个系统的核心,卡通编辑模块和卡通动画模块都以它的生成结果为基础.

(2) 卡通编辑模块.提供给用户的交互式编辑工具,用来生成带有各种表情或夸张效果的卡通.用户可以使用系统预定义的模板产生特定的表情或夸张效果.同时,还提供了一般的编辑功能,编辑结果可存成模板或作为卡通动画模块的关键帧.

(3) 卡通动画模块.生成语音驱动的卡通动画.该模块采用实时的唇形合成算法,根据语音合成卡通的唇形.该算法使用了语言无关的底层语音特征,训练出的模型可以适用于不同语言.



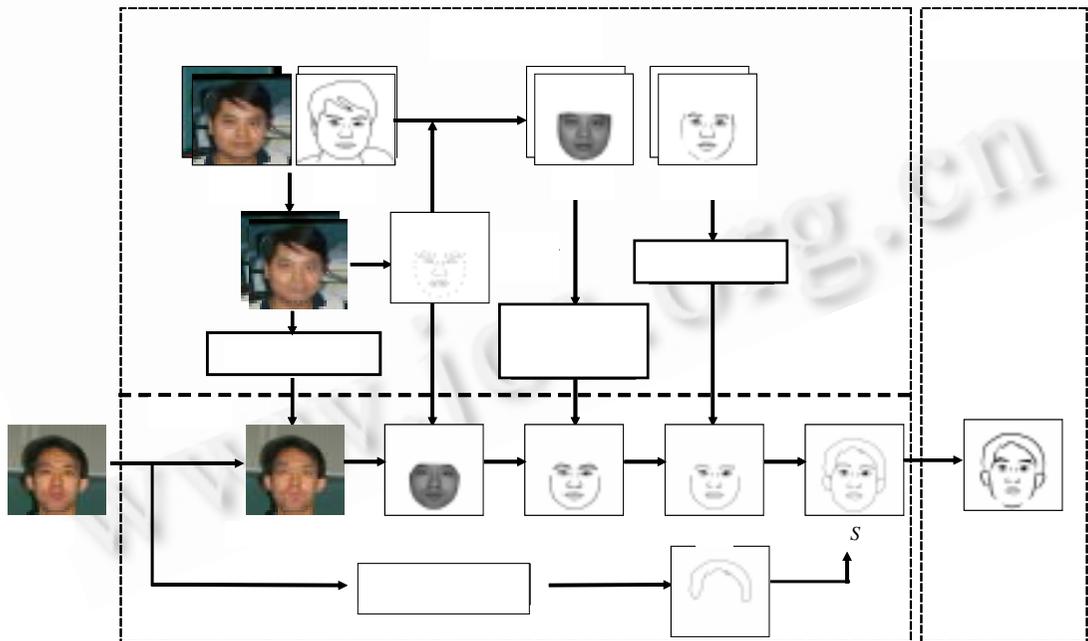
输入图像, 卡通生成模块, 卡通头像, 语音, 卡通编辑模块, 具有夸张表情的模板库, 卡通动画模块, 夸张的卡通, 卡通动画序列.

Fig.1 System architecture

图 1 系统结构

### 3 卡通自动生成

在绘画书籍<sup>[10]</sup>中提到,人脸肖像画有两个关键因素:线条画和线条风格.如图2所示,线条画反映了人脸最基本的视觉特征:简单的线条不但勾画出了人脸的整体结构,而且强调了一些细微的结构特征,如双眼皮等.而线条的风格,使画面具有特定的艺术效果.因此,如图2所示,卡通的生成可以分解为两个步骤:根据输入的人脸图像使用基于样本学习的方法生成矢量化的线条画;绘制出特定笔划风格的卡通画.



线条画生成, 平均脸, 模板匹配, ASM模型, 非参数化采样, 训练, 运行时间, 特征点, 头发轮廓提取, 笔划绘制, 卡通.

Fig.2 Cartoon generation. The approach includes two steps: sketch generation and stroke rendering

图 2 卡通生成框图.算法包含两步:线条画生成和笔划绘制

### 3.1 基于样本的线条画生成

#### 3.1.1 基于样本学习的方法

要从一幅人脸图像生成线条画,本文的系统使用了基于样本学习的方法.如图 2 所示,训练样本集包括了一组人脸图像和画家所画的对应的线条画.本文中,要求训练样本和测试图像都满足如下假设:原始图像都是简单背景下的正面人脸图像(没有帽子和眼镜);线条画未经过夸张,线条和对应的图像特征是对齐的.

在本文的工作中,假定  $S$  和  $I$  之间的统计关系满足以下 3 点假设:

- 局部性假设:线条画上每一点取值  $S(q)$  的概率分布只与原始图像中的对应点周围的一小块邻域  $N_I(q)$  相关.即  $p_q(S(q)|I) = p_q(S(q)|N_I(q))$ .

- 非均匀假设:由于人脸的结构性,上述的条件概率与  $q$  点所处的人脸位置相关.

- 一致性假设:对于同一组训练集中不同人脸的相同位置,上述条件概率是一致的.

考虑到非均匀假设,需要建立图像坐标相对于人脸特征的对立关系.本文的系统根据训练集中的数据,首先得到人脸特征点的平均值,称为平均脸(MeanShape).然后通过图形变形算法使人脸的特征点与平均脸对齐.这样得到的图像,它们的坐标就是对齐过的.不妨令下文中的  $q$  就是人脸特征对齐后的坐标.系统中使用 ASM 模型<sup>[15]</sup>可以自动定位人脸的特征点.

通过训练样本,我们使用了非参数化采样的方法来估计  $p_q(S(q)|N_I(q))$ .然后对这个分布进行积分,得到“期望的线条画” $T'$ ,每一点的取值就是该点出现在线条上的概率.再使用模板匹配就可以得到量化的线条画.

相对于人脸,头发的变化很大,很难保证有一致的统计特性.因此,头发部分没有采用基于样本学习的方法,而是专门设计了提取头发轮廓的算法.

综上所述,如图 2 所示,对输入的人脸图像  $I$ ,我们按如下步骤生成对应的线条画:

##### (1) 人脸部分线条画的生成.

- 人脸特征点定位;
- 几何变换  $I' = G(I)$ ;
- 非参数化采样得到“期望的线条图像” $T'$ ;
- 在  $T'$  上使用模板匹配得到量化的线条化  $S'$ ;
- 使用几何反变换  $S = G^{-1}(S')$ .

##### (2) 提取头发轮廓,生成完整的线条画.

下面给出详细的算法细节.

#### 3.1.2 非参数化采样

非参数化采样方法在纹理合成领域得到了广泛的应用<sup>[16]</sup>.它使用样本来直接构造复杂的统计关系,并从中采样.基于这种思想,针对具体的概率模型,我们采用非均匀的非参数化采样方法来估计局部的条件概率  $p_q(S(q)|N_I(q))$ .

根据假设,对每一点  $q$ ,每个训练样本线条画中的像素点及其在原始图中对应的局部邻域就是  $p_q(S(q)|N_I(q))$  的一个样本.同时,不妨假设统计关系在小的邻域内是稳定的,所以,还可以包含它周围  $l \times l$  邻域内的样本.这种方法还可以消除由 ASM 的定位误差引起的经过几何变换后仍然存在的对齐误差.然后,根据输入的  $N_I(q)$ ,从中选择  $K$  个最近的样本来估计生成的线条画在该点的分布.不妨记它们的集合为  $\Omega(q) = \{S^i(q), N_I^i(q), i = 1, \dots, K\}$ .特别地,在纹理合成领域,生成的点只在样本点的集合中取值,即  $S(q) \in \{S^i(q), i = 1, \dots, K\}$ .

因此,我们定义:

$$p_q(S(q)|N_I(q)) = \sum_{i=1}^K \alpha_i \delta(S(q) - S^i(q)). \quad (1)$$

其中

$$\alpha_i = \frac{1}{Z_q} \exp(-d(N_I(q), N_I^i(q))), \quad (2)$$

式中的  $d(\cdot, \cdot)$  是邻域的距离度量,本文中采用了欧式距离,  $Z_q$  是归一化常数.

对上述的分布求期望,就得到“期望的线条画” $T'$ :

$$T'(q) = \sum_{i=1}^K \alpha_i S^i(q). \quad (3)$$

对于算法中的参数,可以通过实验得到它们的优化值.对于近邻个数  $K$  的选择,实验结果表明,较大的  $K$  会丢失重要的细节,反之,结果会有一些噪声.通过权衡,系统选择近邻个数  $K$  为样本数目的 25%.同时,我们还研究了图像邻域大小对生成结果的影响,最后选择  $7 \times 7$  的邻域.实验表明,在目前的定位精度和对齐算法下,局部搜索窗口的大小  $l$  在一定范围内对结果的影响不是很大.这里我们选择了  $3 \times 3$  的搜索窗口.为了加速采样的过程,与快速纹理合成算法<sup>[17]</sup>一样,本文的系统中也使用了  $3 \times 3$  的图像块作为采样生成的基本单元.

### 3.1.3 模板匹配

模板匹配用于提取矢量化的线条画.为了尽量保持画家的风格,有必要定义一个足够灵活线条画模板.根据训练样本中的线条画,我们定义了一个统一的模型来描述不同人脸的线条画.在本文的系统中,线条画是由一组带有绘制开关的线条组成的.我们记参数化的线条画为  $S = \{L_i : i = 1, 2, \dots, k\}$ , 其中  $k$  是出现在该线条画中所有线条的总数;对于第  $i$  根线条  $L_i$ ,  $L_i = \{c_i, \theta_i\}$ . 其中:

- $c_i$  为开关控制量.  $c_i = 0$  表示这根线条不被绘制,  $c_i = 1$  表示该线条应该被绘制;
- $\theta_i = \{(x_j, y_j) : j = 1, 2, \dots, n_i\}$  为该线条的控制点集,  $n_i$  是控制点的数目.

开/关控制保证了线条画模型的灵活性,可用来表示一些人脸的细节特征,如双眼皮是否存在等.

为了在“期望的线条画”上进行模板匹配,我们首先从训练样本中得到线条画的先验模型.这里,我们假设每一线条间都是统计独立的,建立了一个简化的线条画先验模型:

$$p(S) = \prod_{i=1}^k p(c_i) p(\theta_i), \quad (4)$$

其中,对第  $i$  条线,  $p(c_i)$  是它应当出现的概率,可以用样本中出现的频率来估计.  $p(\theta_i)$  是线条控制点的先验,类似于 ASM 模型<sup>[15]</sup>.我们使用了高斯模型,并使用训练样本估计它的参数.

最后一步就是在得到的“期望线条图”上进行线条画模板匹配.对我们的问题而言,在经过 ASM 的粗定位和几何对齐之后,每根线条的可能解被限制在一个很小的范围内.特别是每条曲线只有少数的控制点,换言之,其参数空间的维数比较低.所以,为了得到每根线条的参数值,我们首先直接根据  $P(\theta_i)$  采样出一组线条,再选择其中最理想的结果,然后通过局部搜索得到邻近的最优解.最后再根据它的似然度和开关控制的先验  $P(c_i)$  来决定每一条线是否应当出现.

### 3.1.4 提取头发轮廓

在进行头发分割时,首先使用训练好的 ASM 模型搜索到大致的头发后轮廓线.然后提取后轮廓内的一小部分区域作为头发灰度的样本.同时,根据前述步骤中得到的人脸特征点位置,也可取人脸区域中的一块为样本.这样即可估计出头发和人脸的灰度分布,从而得到它们的最佳阈值.这样分割后得到的轮廓在边缘上会有一些毛刺,因此采用多边形简化对轮廓进行后处理,去除体积较小的毛刺,同时保持住较大的尖角.

在图 3(b)中,我们给出了提取的头发轮廓与已生成的人脸线条画结合后的结果.

## 3.2 笔划绘制

如图 3 所示,笔划绘制就是采用特定风格的笔触绘制线条画中的线条.每条笔划包括如下属性:

- 宽度:笔划的整体宽度;
- 路径:笔划的骨架,由一组控制点构成.

本系统在绘制笔划时,笔划路径用通过控制点的 Catmull-Rom 曲线表示.系统中使用了宽度渐变的笔划,在笔划路径的两端逐步变细.通常线条画中的每一线条都会绘制成一条笔划.但画家为了增强特定风格的表现力,会把一根线条拆分为若干段笔划.比如,在绘制图 3 中的卡通时,人脸的轮廓线分 3 段绘制.特别是对于头发轮廓,

自动地在很强的拐点处插入断点,可使绘制的结果更加自然.



Fig.3 Generated Cartoon

图 3 生成的卡通

#### 4 卡通编辑模块

卡通编辑模块是提供给用户的图形编辑器,在这里,用户可方便地修改生成的卡通,例如改变其大小、形状或者删除、添加线条等.为了方便用户使用,系统还提供了一系列预先定义的特定表情或夸张的模板,包括微笑、大怒、做鬼脸等.同时,系统还提供了可产生渐变表情状态的编辑工具.图 4 是卡通编辑模块的用户界面示意图.

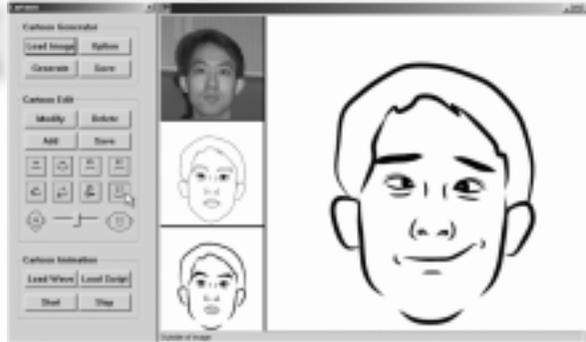


Fig.4 User-Interface of Cartoon editor

图 4 卡通编辑模块的用户界面

在卡通编辑模块中,卡通头像被划分为 7 个部分:脸轮廓、嘴、鼻、左右眼和左右眉毛.对每一部分均可进行以下操作:

- (1) 删除:删除原有的线条.
- (2) 增加:增加新的线条.
- (3) 修改:使用者可以缩放、旋转各部分,还可以拖动笔划骨架的控制点改变线条的形状.

(4) 使用预制的模板:使用者可以选择预先定义的模板来产生具有特定表情和夸张的效果.模板的使用又分为替换模式和改变模式.替换模式直接用模板替换原有的部件.改变模式则纪录了原有部件相应变化的相对值,在这种模式下,系统提供了显著系数来实现渐变效果.

- (5) 保存为模板:使用者可以把编辑的结果保存为模板,保存时,要指定上述的模板类型.

使用这个简单的工具,用户即可方便地设计特定的表情和夸张效果.特别是模板的使用简化了这个过程.在图 5 中给出了使用系统预定义模板编辑表情的结果示例.

#### 5 卡通动画模块

为了让使用者方便地生成动画,我们采用了语音驱动的唇形自动合成算法.本文使用了一个实时的唇形合成算法,它是卡通动画模块的核心.我们没有使用常用的 phoneme-viseme 映射<sup>[13,14]</sup>方法,而是以底层的语音特征作为输入(如语音识别中应用的 MFCC 等).所以,我们的算法同语言无关,即训练出的模型可以适用于不同语言.

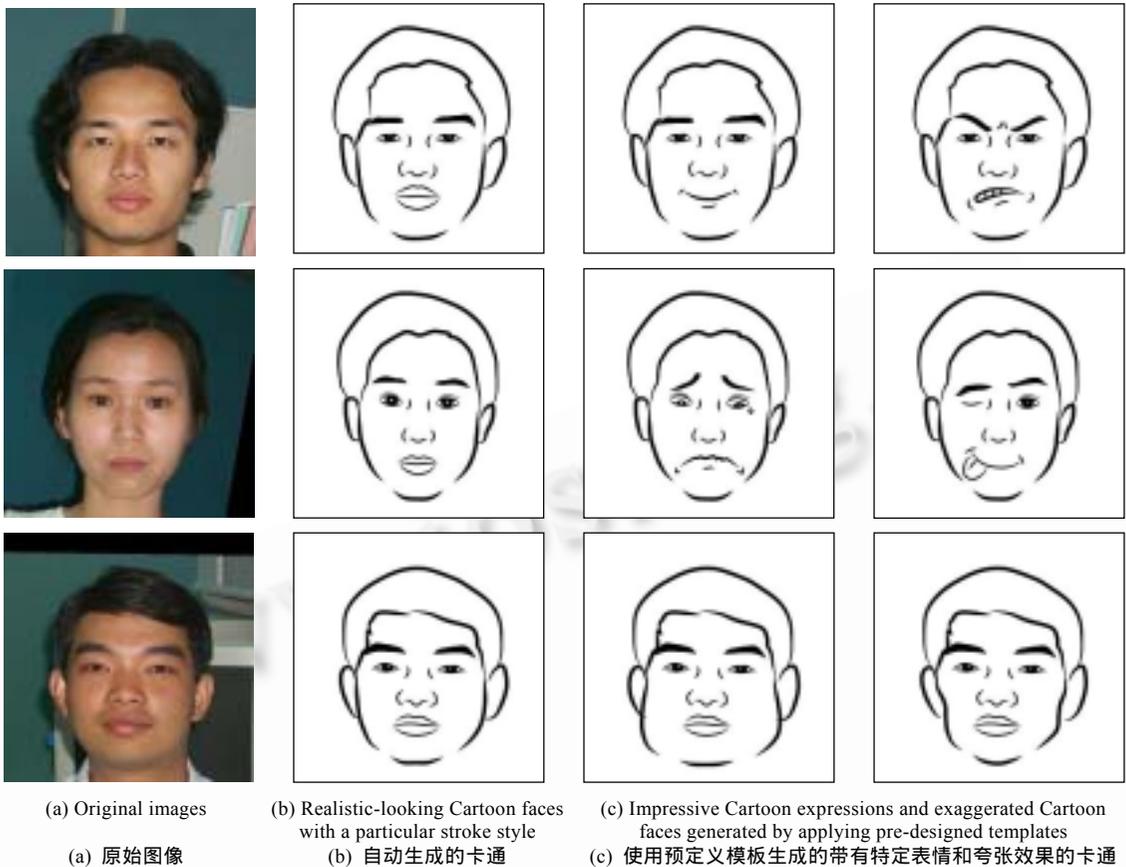


Fig.5 Cartoon faces generated by this system

图 5 系统生成的卡通

5.1 实时唇形合成的模型学习

本文选择了 MFCC(mel-frequency cepstrum coefficients)和 delta coefficients<sup>[18]</sup>等 18 维的语音特征作为输入信号  $\alpha$ . 这些特征可从不同说话人和不同录制环境中鲁棒地提取出来,因而在语音识别领域得到了广泛的使用. 本系统生成的矢量化卡通是由一组二维的控制点表示,把控制唇形的控制点坐标排列成向量,记为  $v$ . 假设它服从于混合高斯分布. 每一个高斯分布的参数包括权值  $\pi_i$ 、均值  $\mu_i$  和协方差矩阵  $R_i$ .

为了从训练样本中得到该混合高斯分布,本文使用了非监督的学习算法<sup>[19]</sup>. 训练数据最终被聚为  $n$  个类(在本系统中  $n=15$ ). 我们称每个类的中心为基本唇形. 所有的唇形都可以用这组基本唇形的线性组合来表示.

给出了这组基本唇形,下一步是要识别出训练序列中每一个唇形的类别. 我们使用了马氏距离作为相似度量,即唇形  $v$  属于第  $j$  类当且仅当:

$$j = \arg \min_k (v - \mu_k)^T R_k^{-1} (v - \mu_k), \quad k = 1, \dots, n. \tag{5}$$

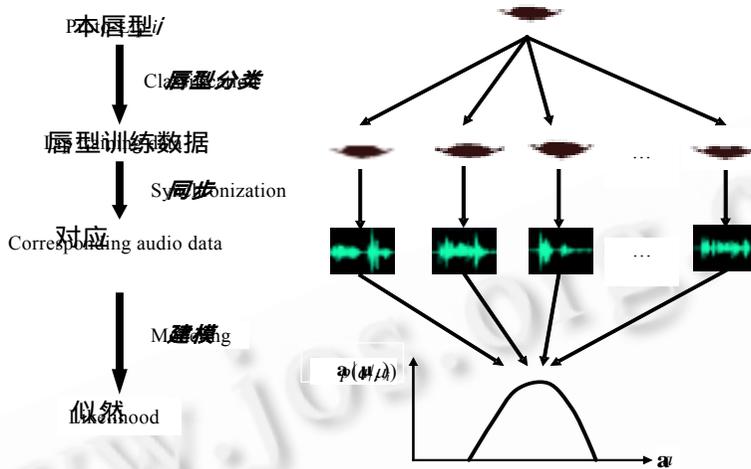
在训练序列中,每一个时刻都有一对语音特征和唇形,因此,我们把语音特征数据也按照对应的唇形类别分类. 同时,假设语音特征的每一维是独立的高斯分布. 这样可以估计出每一类语音特征在每一维的均值  $\alpha_{ij}$  和方差  $\sigma_{ij}$ . 学习过程如图 6 所示.

5.2 唇形合成

如图 7 所示,对于输入的语音序列,首先把它分割成 40 毫秒长的帧. 再将每一帧计算出的底层语音信号特征  $\alpha$  作为系统输入. 因为我们假设语音信号特征的每一维都相互独立,并满足高斯分布,因此,它与第  $i$  个基本唇

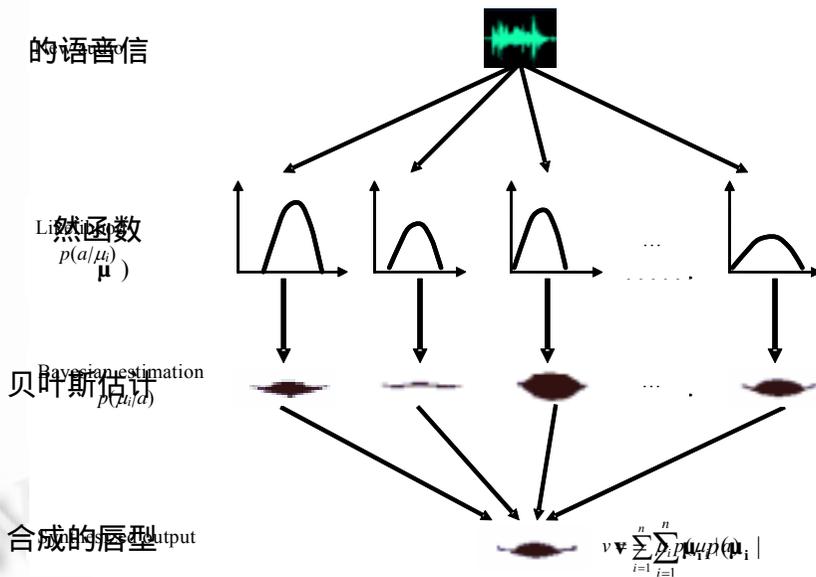
形  $u_i$  的似然值为

$$p(\alpha | \mu_i) = \prod_{j=1}^{18} \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp\left(-\frac{(\alpha_j - \alpha_{ij})^2}{2\sigma_{ij}^2}\right) \quad (6)$$



基本唇型  $i_j$ , 分类, 唇型训练数据, 同步, 对应的语音数据, 建模, 似然函数.

Fig.6 Model training of lip-synching algorithm  
图 6 唇型合成的模型学习算法



新的语音信号, 似然函数, 贝叶斯估计, 合成结果的输出.

Fig.7 Synthesis process of lip-synching algorithm  
图 7 唇型合成算法的合成过程

由贝叶斯公式,对应的属于各类的后验概率为

$$p(\mu_i | \alpha) = \frac{p(\alpha | \mu_i) p(\mu_i)}{\sum_{i=1}^n p(\alpha | \mu_i) p(\mu_i)}, \quad (7)$$

其中  $p(\mu_i) = \pi_i$ , 最终的唇形为

$$\mathbf{v} = \sum_{i=1}^n \mu_i p(\mu_i | \alpha). \quad (8)$$

由于映射误差和噪声的存在,生成的序列中会出现不应当的抖动.算法使用了高斯滤波器,使生成序列在相邻帧之间的过渡更加平滑.

## 6 结果与讨论

图 5 给出了本系统生成的结果示例.图 5(b)中是基于样本方法自动生成的卡通.系统使用的训练样本和输入图像都为  $256 \times 256$ ,样本数目为 40 个,在奔腾 IV1.4Ghz PC 上自动生成的时间是 2 秒.由图 5 可以看出,自动生成的卡通具有特定的艺术风格,同时很好地提取了人脸特征,包括一些细微的特点如双眼皮等.图 5(c)是使用系统预定义的模板生成的各种表情和夸张效果.第 3 行中的胖瘦夸张效果使用渐变编辑方式得到.可以看出,这些生动的结果仍然保持了一定原始人脸的特征.图 8 是系统生成的卡通动画序列.嘴唇的运动由实时的语音驱动的唇形合成算法自动生成.

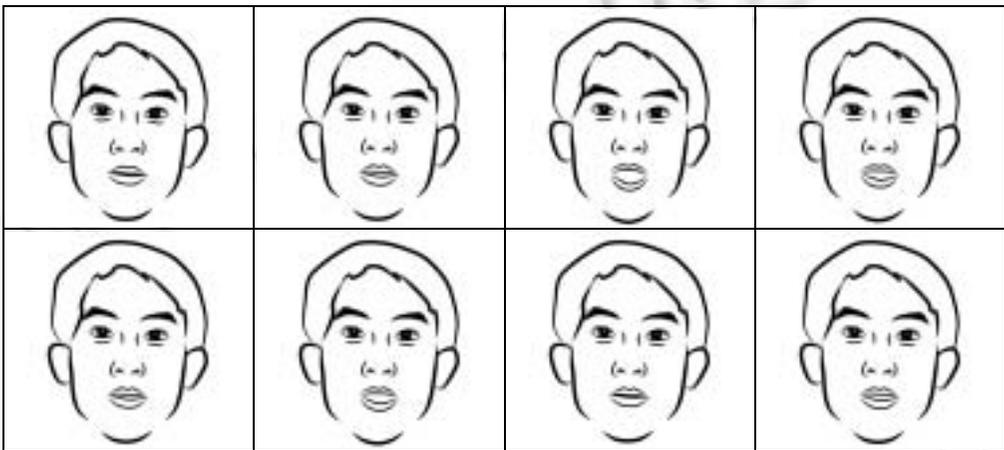


Fig.8 Generated Cartoon animation sequence

图 8 生成的卡通动画序列

本文介绍的卡通系统,可以根据输入的图像自动生成卡通头像.生成的卡通保持了原有的人脸图像的特征,并具有特定的艺术风格.用户可以方便地使用卡通编辑模块生成各种生动表情和夸张效果.系统还可以自动地生成语音驱动的卡通动画.本文的系统可与很多应用相结合,例如电子贺卡、网上聊天.

关于今后的工作,可在以下几个方面进一步改进本文的系统.目前只能自动地生成正面人脸的卡通,可尝试使用三维人脸模型把系统扩展到生成任意位置的卡通.另外,现在是通过用户的交互生成夸张的卡通,如何自动地进行夸张也是一个很有趣的问题.在动画生成方面,目前只能自动合成唇形,在今后的工作中希望能够根据从语音中分析出的感情,自动地生成带有表情变化的动画序列.

## References:

- [1] Hsu, S.C, Lee, I.H.H., Wiseman, N.E. Skeletal strokes. In: UIST'93 Proceedings of the ACM SIGGRAPH and SIGCHI Symposium on User Interface Software and Technology Secrets of the Face. 1993. 197~206.
- [2] Ostromoukhov, V. Digital facial engraving. In: Proceedings of the ACM SIGGRAPH 1999. 1999. 417~424.
- [3] Durand, F. Decoupling strokes and high-level attributes for interactive traditional drawing. In: Proceedings of the Eurographics Rendering Workshop 2001. 2001. 71~82.
- [4] Ruttkay, Z., Noot, H. Animated cartoon faces. In: Proceedings of the 1st International Symposium on Non-Photorealistic Animation and Rendering 2000. 2000. 91~100.
- [5] Flash. <http://www.macromedia.com/software/>.

- [6] Litwinowicz, P.C. Inkwell: a 2.5-d animation system. *Computer Graphics*, 1991,25(4):113~122.
- [7] Brennan, S.E. Caricature generator [MS. Thesis]. Cambridge, MA: MIT Press, 1982.
- [8] Tominaga, M., Fukuoka, S., Murakami, K., *et al.* Facial caricaturing with motion caricaturing in PICASSO system. In: Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics'97. 1997. 30~37.
- [9] <http://www.koshi-lab.sccs.chukyo-u.ac.jp/~fuji/pica2>.
- [10] Edwards, B. The new drawing on the right side of the brain. Harper Collins, 1999.
- [11] Sherstinsky, A., Picard, R.W. M-lattice: a novel non-linear dynamical system and its application to halftoning. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vol 2. 1994. II/565~II/568.
- [12] Wong, E.C. Artistic rendering of portrait photographs [MS. Thesis]. Cornell University, 1999.
- [13] Bregler, C., Covell, M., Slaney, M. Video rewrite: driving visual speech with audio. In: Proceedings of the ACM SIGGRAPH'97. 1997. 353~360.
- [14] Morishima, S., Aizawa, K., Harashima, H. An intelligent facial image coding driven by speech and phoneme. In: Proceedings of the IEEE ICASSP. 1989. 1795-1798
- [15] Cootes, T.F, Taylor, C.J. Statistical models of appearance for computer vision. Technical Report, Manchester: University of Manchester, 2000.
- [16] Efros, A.A., Leung, T.K. Texture synthesis by non-parametric sampling. In: Proceedings of the 7th International Conference on Computer Vision. 1999. 1033~1038.
- [17] Xu, Ying-qing, Guo, Bai-ning, Shum, H. Chaos mosaic: fast and memory efficient texture synthesis. Technical Report, MSR-TR-2000-32, Microsoft Research, 2000.
- [18] Rabiner, L., Juang, B.H. Fundamentals of Speech Recognition. Prentice Hall, 1993.
- [19] Bouman, C.A. Cluster: an unsupervised algorithm for modeling Gaussian mixtures. Software Manual, <http://www.ece.purdue.edu/~bouman>.

## A Personalized Image-Based Cartoon System\*

CHEN Hong<sup>1</sup>, ZHENG Nan-ning<sup>1</sup>, LIANG Lin<sup>2</sup>, LI Yan<sup>2</sup>, XU Ying-qing<sup>2</sup>, Shum Heung-Yeung<sup>2</sup>

<sup>1</sup>(Institute of AI and Robot, Xi'an Jiaotong University, Xi'an 710049, China);

<sup>2</sup>(Microsoft Research Asia, Beijing 100080, China)

E-mail: {hchen,nnzheng}@aiar.xjtu.edu.cn; {i-liang,yli,yqxu,hshum}@microsoft.com

<http://www.aiar.xjtu.edu.cn>; <http://research.microsoft.com/asia/group/research/visualize.asp>

**Abstract:** In this paper, a Cartoon system is presented, which can generate a personalized Cartoon face from an input image. The system is designed for general users. After a realistic-looking Cartoon face automatically generated by this system, attractive Cartoon expressions and animation can be created with little user-interactions. The system consists of three main components, an image-based automatic Cartoon generator, a template-based Cartoon editor, and a speech-driven Cartoon animator. An example-based approach is taken by Cartoon generator to generate the Cartoon face while capturing an artist style. An inhomogeneous non-parametric sampling is employed to learn the statistical relationship between the original image and the drawing. For ease editing, pre-designed expression templates are supplied by Cartoon editor. A real time lip-syncing algorithm is also developed to make Cartoon animation easily. The Cartoon and animation generated by this system have personal style. This is the main advantage of this system over others.

**Key words:** non-photorealistic rendering; example-based learning; non-parametric sampling; lip-syncing

\* Received March 5, 2002; accepted June 18, 2002

Supported by the National Natural Science Foundation of China under Grant No.60024301