

# 基于线索引导的面部运动捕捉与数据虚拟重用方法

张 超 金龙斌 韩 成

(长春理工大学计算机科学技术学院 长春 130022)

**摘 要** 多数无标记面部表情运动捕捉方法仅捕捉面部表情运动的平面位置变化,而无纵深变化描述。针对这一问题,提出了基于线索引导的面部运动捕捉与数据虚拟重用方法。首先,使用单目视觉系统对人脸主体区域进行定位;然后,采用级联回归模型对人脸地标特征逐步求精,通过利用人脸特征的主动地标线索和地标特征的深度线索得到特征点在三维空间上的位置变换关系;最后,借助面部骨骼节点的数据表示实现面部表情运动重构。通过在线的实时面部表情运动捕捉实验效果可以看出,该方法不仅能够实现不同视角下对应地标特征的精确匹配,而且能够较好地将其真实面部运动赋予虚拟角色。

**关键词** 线索引导,面部运动捕捉,级联回归模型,人脸特征

中图分类号 TP391 文献标识码 A

## Method of Facial Motion Capture and Data Virtual Reusing Based on Clue Guided

ZHANG Chao JIN Long-bin HAN Cheng

(School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130022, China)

**Abstract** Most of the unmarked facial expression motion capture methods only capture the planar position changes of facial expression movements, have no description of the depth changes. For this problem, a clue guided based facial motion captures and data virtual reusing method was proposed. Firstly, the method uses a monocular vision system for locating the main body of the face. Then, the face landmark features are refined by the cascade regression model. We can obtain the positional transformation relationship of the feature points in the three-dimensional space by the active landmark cues of the facial features and the depth clues of the landmark features. Finally, the facial skeleton node data are used to achieve facial expression motions' reconstruction. Through the experiment of the online real-time facial expression motion capture, it can be seen that this method can not only achieve the exact match of the corresponding landmark features in different perspectives, but can also better assigns real facial motion to virtual characters.

**Keywords** Clue guided, Facial motion capture, Cascade regression model, Facial feature

## 1 引言

近年来,人体运动捕捉技术<sup>[1]</sup>和面部表情运动捕捉技术<sup>[2-3]</sup>已被广泛地应用在多个领域,利用面部表情运动捕捉技术制作的 3D 科幻影视作品,很好地将人脸面部表情叠加到各类虚拟角色中去,赋予了虚拟角色新的生命,更给广大影视观众带来了震撼的视觉享受。正是如此,面部表情运动捕捉与数据虚拟重用等运动捕捉技术已经成为数字影视摄制技术中必不可少的要素。

现有的人脸特征定位方法主要分为基于显式形状的人脸特征定位方法和基于隐式形状的人脸特征定位方法。基于显式形状的人脸特征定位方法通过学习参数化模型来描述人脸的形状特征,其典型方法有基于活动形状模型<sup>[4]</sup>的方法和基于主成分分析<sup>[5-6]</sup>的方法等;基于隐式形状的人脸特征定位方法则直接从输入图像进行人脸关键特征的位置回归<sup>[7]</sup>,其典型方法有基于深度学习模型<sup>[8-9]</sup>的方法和基于级联回归模

型<sup>[10-11]</sup>的方法等。

针对无标记面部表情运动捕捉方法中存在的位置描述限制问题,本文提出了一种基于线索引导的面部运动捕捉方法。该方法在无标记面部表情运动捕捉实现过程中,将在人脸主体区域定位的基础上,根据面部表情的初始地标特征采用级联回归模型的人脸特征定位方法对人脸地标特征逐步求精,最后利用双目视觉系统获取主动地标特征的深度线索信息,以描述面部表情骨骼节点的三维位置变化。

## 2 级联回归模型

传统的人脸特征点定位算法需要手工对训练样本集进行定位,其对复杂且大量的人脸图像是不可靠的,而基于级联回归模型的人脸特征定位则利用了若干级联回归器对人脸特征点之间的约束,来不断更新级联回归器中的参数<sup>[12]</sup>。由于级联回归模型采用局部描述子对人脸特征点进行定位,其不易受到外界环境因素的影响,且级联回归模型对复杂的非线性

本文受吉林省科技攻关计划项目(20170203003GX),吉林省科技攻关计划项目(20170203004GX)资助。

张 超(1985—),男,博士,讲师,CCF 会员,主要研究方向为图像处理、计算机视觉、增强现实等,E-mail: zhangchao@cust.edu.cn;金龙斌(1991—),男,硕士生,主要研究方向为虚拟现实与多媒体技术;韩 成(1978—),男,博士,副教授,主要研究方向为数字媒体与虚拟现实,E-mail: hancheng@cust.edu.cn(通信作者)。

函数有较强的映射能力,因此级联回归模型成为了研究的热点。

基于级联回归模型的人脸特征点定位,首先通过人脸检测器对输入图像进行检测来获得初始人脸形状  $I_0^s$ ,通过回归器进行回归残差求解,再将多个级联回归器  $M_R = \{R_1, R_2, R_3, \dots, R_T\}$  迭代拟合求解最优的回归残差。通过式(1)将人脸形状  $I_{i-1}^s$  更新为  $I_i^s$ ,其中  $R_i(I, I_{i-1}^s)$  表示弱回归器  $R_i$  对图像  $I$  求解的人脸形状  $I_{i-1}^s$  进行更新的回归残差。将第  $i$  个回归模型中的回归器  $R_i(I, I_{i-1}^s)$  描述为式(2)的形式,其中  $W_i$  为回归器的映射矩阵,  $b_i$  为回归器映射函数的截距。

$$I_i^s = I_{i-1}^s + R_i(I, I_{i-1}^s) \quad (1)$$

$$R_i(I, I_{i-1}^s) = \{W_i, b_i\} \quad (2)$$

为了进行拟合求解回归参数,对目标函数式(3)进行最小二乘法优化求解,其中  $I_i^{s*}$  表示真实的人脸形状,而  $I_i^s$  为迭代过程中求解的人脸形状,  $\lambda$  为防止过拟合优化的控制权重系数,  $f(I, I_i^s)$  为级联回归模型的映射函数,  $f(I, I_i^s)$  是与人脸形状特征相关的函数,  $W_i$  和  $b_i$  为级联回归模型迭代求解的回归参数。

$$\frac{1}{T} \arg \min_{w_i} \sum_{i=1}^T \| I_i^{s*} - I_i^s - R_i(I, I_i^s) \|_2^2 + \lambda \| W_i \|_2^2 =$$

$$\frac{1}{T} \arg \min_{w_i} \sum_{i=1}^T \| I_i^{s*} - I_i^s - W_i f(I, I_i^s) - b_i \|_2^2 + \lambda \| W_i \|_2^2 \quad (3)$$

通过最小二乘法优化目标函数式(3),求解得到的回归参数为:

$$\begin{bmatrix} b_i^T \\ W_i^T \end{bmatrix} = (F_i F_i^T + \lambda E_n)^{-1} F_i I_M^s \quad (4)$$

其中,  $F_i$  为在求解过程中利用级联回归模型中映射函数构建的矩阵,  $E_n$  为单位对角矩阵,而  $I_M^s$  则为初始人脸形状特征和真实人脸形状特征的差值。求解完第  $i$  个回归器的回归参数后,即可利用式(5)对第  $i+1$  次的人脸形状特征进行更新。

$$I_{i+1}^s \leftarrow I_i^s + W_i f(I, I_i^s) + b_i \quad (5)$$

### 3 基于线索引导的面部运动捕捉方法

由于无标记面部表情运动捕捉需要使用人脸地标特征来描述面部表情运动情况,因此人脸地标特征提取结果的精度将直接影响面部表情运动捕捉的有效程度。因为单目视觉

$$\begin{cases} m(x, y) = \sqrt{(I(x+1, y) - I(x-1, y))^2 + (I(x, y+1) - I(x, y-1))^2} \\ \theta(x, y) = \arctan((I(x, y-1) - I(x, y+1)) / (I(x+1, y) - I(x-1, y))) \end{cases} \quad (6)$$

对每个单元块内的梯度信息进行梯度方向直方图统计,也就是对计算的梯度方向在空间中分割成 8 个角度区间,对每个单元块内相同的梯度方向进行加权投影统计,其中加权投影采用梯度幅度值进行加权,统计完成后得到单元块的 8 维梯度方向直方图。将移动窗口中 4 个单元块的 8 维梯度方向直方图进行联合统计,得到一个 32 维的梯度方向直方图特征信息。当移动窗口对整幅图像都遍历完成后,级联所有移动窗口遍历的 32 维梯度方向直方图特征,即得到该图像的 HOG 特征向量。

由于 AdaBoost 算法是一种自适应训练学习算法,其自适应地对若干弱分类器进行更新,权重值构建成为一个强分类器。因此在得到图像的 HOG 特征向量后即可利用 AdaBoost 算法对其进行人脸区域定位,AdaBoost 算法将每个 HOG 特征对应一个弱分类器。

系统只能描述人脸地标特征的平面运动情况,而无法描述人脸地标特征的三维运动情况,所以在进行面部表情运动捕捉时,将使用双目视觉系统来捕捉左、右视角的人脸图像。为此,提出了一种基于线索引导的面部运动捕捉方法,该方法利用级联回归模型方式对人脸地标特征进行精确定位,然后根据左、右视角人脸地标特征的匹配关系计算深度值,从而根据地标线索信息和深度线索信息来描述人脸地标特征在三维空间中的运动状态。

基于线索引导的面部运动捕捉过程如图 1 所示,首先分别对左、右视角人脸图像进行区域定位和地标特征提取;然后根据左、右视角人脸地标特征的匹配关系估算各个地标特征的三维坐标位置;最后对人脸地标特征点的运动情况进行描述,并将该面部表情传递给三维虚拟模型,以实现真实面部表情的虚拟化。

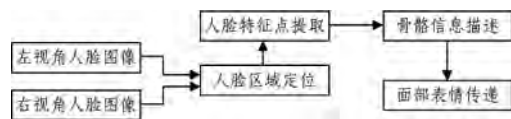


图 1 基于线索引导的面部运动捕捉过程

#### 3.1 主体人脸区域检测

基于 AdaBoost 的人脸检测定位算法是一种训练学习的过程,首先需从人脸样本 5 图像中提取出特征值,使用人脸特征值中具有较强分类的特征信息来构建 AdaBoost 算法的弱分类器,Ada-Boost 算法通过迭代更新弱分类器的权值,最后将弱分类器结合成一个能够实时对图像进行人脸定位的强分类器。由于 HOG 算法凭借着梯度特征信息描述子进行人脸特征提取时能够不被外界环境光照等干扰,因此首先通过 HOG 算法检测人脸特征信息,再结合 AdaBoost 算法对检测到的人脸特征信息进行人脸定位。

HOG 算法的具体实现步骤如下:定义一个移动网格窗口,该窗口由以特征点为中心的  $8 \times 8$  像素邻域大小构成,其中窗口移动的步长可根据图像大小进行定义。对移动窗口进行分单元块处理,即将移动网格窗口分割成 4 个单元块,每个单元块的大小都是  $4 \times 4$  像素。采用中心对称梯度算子对像素的梯度值求解,即利用式(6)求解每个单元块中像素的梯度信息。其中,  $I(x, y)$  为移动窗口中点  $(x, y)$  处的像素值,而  $m(x, y), \theta(x, y)$  分别为该像素的梯度幅值和梯度方向。

AdaBoost 算法的步骤如下:设 AdaBoost 算法的训练集有  $n$  个样本,分别为  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ ,其中为了区分样本是否为人脸,用  $y_i = \{1, 0\}$  表示真、假人脸样本,而用  $f_m(x_i)$  表示人脸图像样本  $x_i$  的  $m$  个特征向量信息。为每个  $f_m(x_i)$  构建对应的弱分类器,如式(7)所示:

$$h_j(x) = \begin{cases} 1, & p_j f_j(x) < p_j \theta_j \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

其中,  $p_j$  一般取值为  $+1$  或  $-1$  并决定不等号的方向,而  $\theta_j$  表示该弱分类器的阈值。对每个输入的人脸样本进行初始权值设定,如果样本有  $l$  个真值人脸样本,则将真值人脸样本的权值设置为  $w_{1,i} = 1/2l$ ,而将非人脸样本的权值设置为  $w_{1,i} = 1/2(n-l)$ 。

AdaBoost 算法一共训练  $T$  次,对每轮  $t=1, 2, \dots, T$  训练都进行以下操作:首先利用式(8)对权值进行归一化处理,

接着对每个特征  $j$  进行弱分类器的训练学习, 通过  $\epsilon_{t,i} = \sum_{i=1}^n \omega_{t,i} |h_{t,j}(x_i) - y_i|$  对训练特征  $j$  进行错误率计算, 最后提取错误率最小的弱分类器。利用  $\omega_{t+1,i} = \omega_{t,i} \times (\epsilon_i / ((1 - \epsilon_i)^{1-\epsilon} + \epsilon_i))$  对  $n$  个样本的权值进行更新, 如果样本被正确分类则  $\epsilon_i = 0$ , 否则  $\epsilon_i = 1$ 。训练完成后, 将所有的弱分类器组合成一个强分类器, 如式(9)所示, 其中  $\delta_t = \log((1 - \epsilon_t) / \epsilon_t)$ 。

$$\omega_{t,i} = \omega_{t,i} / \sum_{i=1}^n \omega_{t,i} \quad (8)$$

$$h(x) = \begin{cases} 1, & \sum_{t=1}^T \delta_t h_t \geq \frac{1}{2} \sum_{t=1}^T \delta_t \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

### 3.2 多尺度人脸特征点提取

在一幅面部表情捕捉图像中, 利用 3.1 节描述的人脸区域定位方法可能会得到多组人脸定位特征, 这里将只选取所占区域最大的一组人脸特征作为当前有效的人脸区域, 该人脸区域记为  $F^*$ 。在人脸区域  $F^*$  中, 首先使用主动形状模型得到初始地标特征, 然后利用级联回归模型对初始地标特征进行精确定位, 从而得到原始尺度下的人脸地标特征的位置。

由于受到光照、噪声等影响, 人脸地标特征的检测结果可能会出现静帧抖动的现象, 本文利用多尺度人脸地标特征检测的方式来消除抖动。经过多尺度地标特征检测后, 使用各个尺度下相同地标特征位置信息的均值来表达该地标特征的位置信息。

### 3.3 面部表情运动重构

对于每一时刻的面部表情运动来说, 仅仅使用当前帧的人脸地标特征将无法准确描述其表情状态, 因为任意时刻的面部表情运动都是相对而言的。从本质上来说, 面部表情运动捕捉就是获取当前时刻人脸地标特征与初始人脸地标特征的相对位移关系。

由于无标记面部表情运动捕捉过程中捕获的是真实人脸的地标特征运动情况, 因此将捕获到的地标特征直接赋予三维虚拟模型有可能会发生表情不明显或表情错位等情况。为了使捕获到的地标特征能够适用于多数三维虚拟模型, 需要以特征位置归一化的形式使用面部表情运动数据。

## 4 实验结果及数据分析

为验证所提算法的可行性、有效性和鲁棒性, 使用多组测试图像在实验结果和数理分析过程中进行了必要的定性与定量分析。算法验证过程使用的计算硬件环境为安装 OS X ElCapitan 10.11 操作系统的 MacBook Pro 计算机, 其硬件配置为: 2.5 GHz Intel Core i7 处理器、16 GB 1600 MHz DDR3 内存、Intel Iris Pro 1536 MB 图形卡。将算法的实现过程分为两个阶段: 人脸区域定位及多尺度特征点提取、面部运动捕捉数据绑定。在实验过程中所使用的验证图像来自于标准测试数据集或特殊要求图像。

### 4.1 人脸区域定位及多尺度特征点提取

本文所提出的无标记面部表情运动捕捉方法中, 首先根据 3.1 节中的人脸区域定位方法确定人脸所在的方形区域, 然后根据确定的人脸有效区域对人脸地标特征进行提取来描述当前面部表情状态。为了验证对人脸区域定位及人脸地标特征点提取的有效性和准确性, 本文进行了多组面部表情图像中人脸区域定位和地标特征提取的检测。图 2 给出了人脸区域检测及地标特征提取效果图, 其包含了 5 组人脸测试图像, 其中: 第一列为面部表情的人脸区域定位; 第二列为基于

主动形状模型的人脸地标特征提取结果; 第三列为基于级联回归模型的人脸地标特征提取结果。



(a) 人脸区域定位 (b) 主动形状模型 (c) 级联回归模型

图 2 人脸区域检测及地标特征提取

从图 2 中人脸地标特征的提取结果可以看出, 在进行人脸地标特征提取时, 仅使用主动形状模型方式得到的人脸地标特征会存在较大的误差, 例如: 最后一组测试图像中使用主动形状模型方式获取的地标特征在嘴角处出现明显偏差; 而使用级联回归模型方式得到的人脸地标特征则能够更为精准地描述人脸特征点位置。因为无标记面部表情运动捕捉严重依赖于人脸地标特征的提取精度, 所以使用级联回归模型的方式来进行人脸地标特征提取具有重要的作用。

由于受到光照、噪声、图像质量等影响, 提取到的人脸地标特征极易出现静帧抖动的现象, 如果对面部表情图像进行必要的缩放操作则会在一定程度上将细微的抖动情况过滤掉。根据不同的光照等条件, 最小可识别的面部表情图像尺寸可能会发生变化, 但在一般情况下, 面部表情图像的尺寸达到宽、高 100 像素左右即可。在对标准图像进行定位与特征提取时, 均假设摄像机与该头部具有相对的刚性不变性, 这样就可以在初始帧的时刻, 根据检测到的人脸区域所占比率得到帧图像的最小可用尺寸, 这样不仅能够保持原有特征点坐标位置的不变性, 还很大程度地减少了运算时间, 提高了实时性。

实验过程中, 原始面部表情图像的宽、高均为 512 像素。表 1 列出了不同尺度特征偏移量和运行时间的对比实验结果, 其中分别给出了面部表情图像被压缩成宽、高 256 像素和宽、高 128 像素尺度下与原始面部表情图像检测值的对比情况。表 1 中, 最小偏差、最大偏差、平均偏差均为当前尺度特征信息与原始尺度特征信息的差值情况。从表 1 中可以看出, 当前尺度人脸图像的检测速度与其下一尺度人脸图像的检测速度基本是 4 倍的关系, 而且检测到的人脸地标特征点并没有出现严重的坐标偏移问题, 因此对小尺度人脸图像进行定位与特征提取完全能够保持原始的面部表情形态且提高了实时性。

表1 不同尺度特征偏移量和运行时间的对比实验分析

尺度空间	最小偏差(pixel)	最大偏差(pixel)	平均偏差(pixel)	原始尺度检测时间/ms	当前尺度检测时间/ms
第二组(256像素)	0	6.4031	1.8709	3.05058	0.793477
第二组(128像素)	0	8.0623	2.7260	3.05058	0.209991
第三组(256像素)	0	10.1980	2.6675	2.85511	0.731844
第三组(128像素)	0	7.0711	2.6080	2.85511	0.199125
第四组(256像素)	0	6.4031	2.7500	2.77207	0.747765
第四组(128像素)	1	19.6469	4.5068	2.77207	0.194472
第五组(256像素)	0	7.6158	2.6750	2.88879	0.745115
第五组(128像素)	1	13.1529	4.3810	2.88879	0.216138

#### 4.2 面部运动捕捉数据绑定

通过使用人脸区域定位及多尺度特征点提取方法<sup>[13]</sup>,可以得到较为精准的面部表情地标特征,它保证了面部表情运动捕捉的数据有效性。在进行面部表情运动捕捉时,使用双目视觉系统对面部表情运动情况进行采集。

图3给出了面部表情特征提取及绑定效果,第一列为左视角面部表情特征的提取结果,第二列为右视角面部表情特征的提取结果,第三列为面部表情与虚拟人物头部骨骼的绑定效果图。其中,对图3中的真实人脸进行模糊化处理以避免侵犯其肖像权。从图3中可以看出,左、右人脸图像的特征提取结果较为可靠,虚拟人物三维头部骨骼的表情绑定结果与真实人物所做的表情动作完全一致,面部表情绑定结果较为理想。



图3 面部表情特征提取及绑定效果

**结束语** 本文重点研究无标记面部表情运动捕捉技术和面部表情运动数据的虚拟重用方法,通过使用面部表情运动数据的虚拟重用方法,可将真实人脸的面部表情赋予虚拟角色。针对无标记面部表情运动捕捉方法中存在的位置描述限制问题,本文提出了一种基于线索引导的面部运动捕捉方法,该方法由主体人脸区域检测、多尺度人脸特征点提取和面部表情运动重构3个步骤组成。基于线索引导的面部运动捕捉方法在对人脸主体区域定位的基础上,根据面部表情的初始地标特征采用级联回归模型的方式对人脸地标特征逐步求精。从人脸地标特征提取的对比实验结果可以看出,未采用级联回归模型方式提取的人脸地标特征会出现严重失真现象,而使用级联回归模型方式提取的人脸地标特征能够较好地表征人脸表情特征。由于待识别图像的尺寸不同,将直接影响面部表情特征的检测时间,因此本文对不同尺度图像的地标特征偏移量和检测时间进行了对比分析。从数据分析结果可以得出,多尺度人脸地标特征的提取能够在不同尺度条件下仍然保持面部表情形态基本不变,并且大幅降低了面部

表情运动捕捉的检测时间,而且在一定程度上消除了人脸地标特征抖动的情况。通过实验结果可以看到,本文提出的基于线索引导的面部运动捕捉方法能够有效地实现虚拟模型面部表情运动的重构。

#### 参考文献

- [1] CHAN J C P, LEUNG H, TANG J K T, et al. A Virtual Reality Dance Training System Using Motion Capture Technology[J]. IEEE Transactions on Learning Technologies, 2011, 4(2): 187-195.
- [2] DARUJATI C, HARIADI M. Facial motion capture with 3D active appearance models[C]// International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering. IEEE, 2013: 59-64.
- [3] 周芳芳, 赵颖, 杨斌. 拟人 Agent 面部复杂运动合成的研究与实现[J]. 计算机科学, 2012, 39(5): 195-197.
- [4] COOTES T F, IONITA M C, LINDNER C, et al. Robust and accurate shape model fitting using random forest regression voting[C]// European Conference on Computer Vision. Springer Berlin Heidelberg, 2012: 278-291.
- [5] LIN S D, LIU B F, LIN J H. Combining speeded-up robust features with principal component analysis in face recognition system[J]. International Journal of Innovative Computing Information & Control Ijicic, 2012, 8(12): 8545-8558.
- [6] 王绍宇. 基于小波和 ICA 的面部特征定位[J]. 计算机科学, 2006, 33(9): 199-200.
- [7] TZIMIROPOULOS G. Project-Out Cascaded Regression with an application to face alignment[C]// Computer Vision and Pattern Recognition. IEEE, 2015: 3659-3667.
- [8] TU C T, LIEN J J. Automatic location of facial feature points and synthesis of facial sketches using direct combined model[J]. IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society, 2010, 40(4): 1158-1169.
- [9] XIONG X, TORRE F D L. Supervised Descent Method and Its Applications to Face Alignment[J]. Computer Vision & Pattern Recognition, 2013, 9(4): 532-539.
- [10] SUN Y, WANG X G, TANG X O. Deep Convolutional Network Cascade for Facial Point Detection[C]// Computer Vision and Pattern Recognition. IEEE, 2013: 3476-3483.
- [11] LI Y Q, WANG S F, ZHAO Y P, et al. Simultaneous facial feature tracking and facial expression recognition[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2013, 22(7): 2559-2573.
- [12] 贾静平, 覃亦华. 基于深度学习的视觉跟踪算法研究综述[J]. 计算机科学, 2017, 44(s1): 19-23.
- [13] 牛耕田, 王昌明, 孟红波. 基于多尺度稀疏表示的面部疲劳识别[J]. 计算机科学, 2016, 43(8): 282-285.