基于高斯过程的表情动作单元跟踪技术

王 磊[1,3],邹北骥[2,3],彭小宁[2,3],潘丽丽[1]
(1.湖南大学计算机与通信学院,湖南长沙 410082; 2.中南大学信息科学与工程学院,湖南长沙 410083;
3.浙江大学计算机辅助设计与图形学国家重点实验室, 浙江杭州 310058)

摘 要: 在表情动作单元的跟踪中有两个常见问题:一是跟踪结果有小幅而频繁的抖动;二是跟踪过程会产生难以检测的误差.针对这两个问题,本文提出了一种基于高斯过程和粒子滤波的表情动作单元跟踪技术.实验结果表明本文算法比传统的梯度优化和粒子滤波法具有更好的平滑性和跟踪精度,而精度的优势在头部有偏转的情况下尤为突出.

关键词: 表情动作单元跟踪; 梯度优化; 粒子滤波; 高斯过程
中图分类号: TP391.41 **文献标识码:** A **文章编号:** 0372-2112 (2007) 11-2087-05

Facial Tracking by Gaussian Process

WANG Lei[1,3], ZOU Bei-ji[2,3], PENG Xiao-ning[2,3], PAN Li-li[1]
(1. *School of Computer and Communication, Hunan University, Changsha, Hunan* 410082, *China*;
2. *School of Information Science & Engineering, Central South University, Changsha, Hunan* 410083, *China*;
3. *State Key Laboratory of CAD & CG, Zhejiang University, Hangzhou, Zhejiang* 310058, *China*)

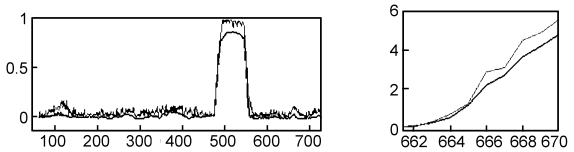
Abstract: Facial Action Units (FAU) tracking is a hard problem for the rigid and non-rigid transformations of human face. The constantly trembling in the tracking result and the tracking failures caused by the absence of constraint remain open problems. This paper presents a novel method to attack these problems by combining Gaussian Process and Particle Filtering. Gradient-based method and particle filtering based method are compared with our method and the experiment results are encouraging.

Key words: FAU tracking; gradient optimization; particle filtering; Gaussian process

1 引言

表情是人类传递情感的重要途径.心理学家归纳出喜悦、惊讶、恐惧、厌恶等基本表情,并将每种基本表情解析为闭左眼、扬左眉、张嘴巴等数十个表情动作单元的组合(Facial Action Units,简称 FAU)[1].本文研究从视频中实时捕捉 FAU 变化的方法.这项技术对于智能人机交互系统和数字娱乐产业具有重要的意义.

已有的 FAU 跟踪技术存在平滑和约束性两方面问题.平滑性问题体现为跟踪结果不够稳定,跳跃感强.比如图 1(*a*)中的曲线是对一段视频中嘴部 FAU 的跟踪结果.其中细实线是由基于粒子滤波的方法[2,3]产生的,粗实线是由本文算法产生的.两条曲线的基本走势相同,但前者的抖动明显要剧烈些.进一步对它们进行频域分析可以得到图 1(*b*),该图显示了跟踪结果经过傅立叶变换后的频率曲线,细实线的高频分量确实高于粗实线.跳跃感强的 FAU 跟踪结果如果应用到三维虚拟表情方面会使合成的表情呈现持续的小幅变化,极不自然;如果应用到表情识别方面会引入额外的状态变迁噪声,降低表情识别率.



(*a*) 嘴部 AU 的跟踪结果曲线 (*b*)跟踪结果的频率分析曲线

图 1 跟踪平滑性问题示例.其中细实线是由一种基于粒子滤波器的跟踪方法产生的,粗实线是由本文提出的方法产生的

FAU 跟踪的另一个问题是难以对跟踪结果实施有效约束.比如图 2 所示的一个瞬时跟踪结果.其中图 2(*a*)显示的是某时刻的人脸视频图像;图 2(*b*)显示的是该时刻的跟踪结果,以三维人脸网格表示.图 2(*c*)是依据跟踪结果合成的三维虚拟人脸表情.在图 2(*b*)和(*c*)中嘴角呈拉伸并下沉状,而图 2(*a*)中的真实人脸却无此表情.虽然这种误差在视觉上很明显,但在数值计算中却往往难以被察觉.这是由于追求实时性的需要,用于跟踪计算的视频图像往往只有较低的分辨率(长宽≤350 像素),人脸区域就更小(一般在 200 像素的尺度以内).在这有限的图像信息中包含了头部姿态、表情、

收稿日期:2007-01-31;修回日期:2007-05-30
基金项目:国家自然科学基金(No.60573079;No.60673093)

光照和遮挡等变化和干扰因素,所以视觉上明显的跟踪误差在数值上却极易淹没在干扰和噪声中不被发现.图 2(*a*)中的干扰因素体现为头部偏转和嘴部阴影,在它们的影响下产生了图 2(*b*)所示的跟踪结果,该结果是应用粒子滤波法所得到的最佳数值解.可见,数值上最优的解未必能如实地反映真实情况.因此对跟踪结果进行合理性约束就显得非常必要了.

(*a*)视频图像 (*b*)跟踪结果 (*c*)3D 合成表情

图 2 FAU 跟踪误差示例

针对上述两个问题本文在第 2 节提出了一种解决方法.其主要思路是对 FAU 用高斯过程法(Gaussian Process,GP)[4~6]进行非线性降维,并在低维隐变量空间中用粒子滤波法进行跟踪计算.维度的降低可以减少粒子采样的离散程度,进而提高跟踪平滑性;而高斯过程的先验概率分布可以帮助我们估计采样粒子与训练样本的相似度,并通过剔除相似度低的粒子来实施跟踪约束.本文的主要贡献是:

(1)将粒子滤波法置于高斯过程框架下进行表情动作单元跟踪;

(2)利用高斯过程的降维效果和降维中的映射相邻性[6]提高粒子滤波法的跟踪平滑度.

(3)利用高斯过程在隐变量空间的先验概率分布对跟踪结果实施有效约束.

虽然利用高斯过程法对物体进行跟踪并非本文的首创[7~9],但本文却首次将高斯过程与粒子滤波法结合起来对人脸这种非刚性物体进行跟踪.考虑到头部的刚性运动随意性很强(比如人们的注视方向可能因为目标的位置变化而朝向任意方向),而且头部姿态与表情变化是彼此独立的,所以本文把头部姿态参数与 FAU 参数分开进行跟踪.对于前者,我们采用梯度优化算法对未经降维的参数进行跟踪;对于后者,我们采用粒子滤波算法在低维隐变量空间中进行采样跟踪.此外为了加强对跟踪结果的约束,本文在 FAU 跟踪中除了用图像特征似然值对粒子权重进行调整外,还利用了高斯隐变量空间的先验概率分布.第 3 节的实验结果表明新方法相对于传统的粒子滤波或梯度优化法具有较好的跟踪平滑性,而且通过对跟踪结果进行先验约束,跟踪错误也显著减少.

2 算法描述

2.1 基于高斯过程的 FAU 概率模型

表情动作单元(FAU)的变化不是彼此独立的,而是在人类情绪等因素的驱动下按照某种规则协同运动的.驱动 FAU 变化的因素很多,比如人的情绪、习惯、年龄、性别、脸形胖瘦等.这些因素都可能引发 FAU 变化的差异性.假设 FAU 的取值为 Y,驱动其变化的因素为 X,通过回归分析我们可以建立 X 与 Y 之间的映射关系,如式(1).

$$\mu(x) = y = \sum_{h=1}^{H} \omega_h \varphi_h(x) \tag{1}$$

其中 Φ 是径向基函数[10],h 是基函数的序号,ω 是基函数的权值,$\varphi_h = \exp\left(-\frac{(x - c_h)^2}{zr^2} \right)$.

由于式(1)中的 X 是未知的,所以我们先用主元分析法估计 X 的取值,然后再对回归参数和 X 进行优化求解.尽管 X 无法与"情绪、习惯、年龄"等因素直接联系起来,但它提供了 FAU 的低维表示方法,使 FAU 的跟踪等价为对 X 的跟踪.这可以大幅降低计算量并改善跟踪精度和平滑性.然而式(1)中的待定参数很多(比如 H 个权值 ω 和 X),因此参数求解的难度较高而且容易过度拟合(overfit).我们采用"高斯过程法"的思路来解决这个问题.假设 $\omega \sim N(0, \delta_\omega^2)$,这样 H 个 ω 便退化为一个方差变量,规则函数 μ 转化为随机过程.由于 FAU 的取值 y 是 ω 的线性组合,根据高斯分布的线性叠加性,FAU 的分布也呈高斯分布,即:

$$P(y) \sim N(0, \delta_\omega^2 \varphi(x)\varphi(x')) \tag{2}$$

其中 $\delta_\omega^2 \varphi(x)\varphi(x')$ 形成协方差矩阵,记为 K,它的第 i 行 j 列元素 $k_{ij} = \delta_\omega^2 \sum_h \varphi_h(x_i)\varphi_h(x_j)$.鉴于 $P(y)$ 的均值为 0,我们可以把 Y 看作是减去了均值的 FAU.

虽然式(2)只揭示了 FAU 的先验概率分布 $P(y)$,但结合贝叶斯定理我们可以实现从 X 到 FAU 的回归映射.假设我们有一组样本 $D = \{x_i, y_i \mid i = 1,2,\cdots,N\}$,其中 N 是样本个数,x 可由上文描述的方法求得.我们希望求出测量值 x_t 对应的 FAU 取值 y_t,这等价于求解条件概率 $P(y_t \mid D)$ 的数学期望.根据贝叶斯定理和逆矩分解公式(partitioned inverse equation)[5]它可以展开为:

$$P(y_t \mid y_D) = \frac{1}{Z} \exp\left[-\frac{(y_t - \tilde{y})^2}{2\tilde{\sigma}^2} \right] \tag{3}$$

其中

$$\tilde{y} = \boldsymbol{k}^{\mathrm{T}} \boldsymbol{K}_N^{-1} \boldsymbol{y}_D \tag{4}$$

$$\tilde{\sigma}^2 = k - \boldsymbol{k}^{\mathrm{T}} \boldsymbol{K}_N^{-1} \boldsymbol{k} \tag{5}$$

其中 $\boldsymbol{K}_N$ 是样本 D 的协方差矩阵,$\boldsymbol{k}$ 为样本 D 和测量值的协方差矩阵 $\boldsymbol{K}_{N+1}$ 最后一列的前 N 个元素,k 是 $\boldsymbol{K}_{N+1}$ 的右下角元素.由于 Y 是减去了均值的 FAU,所以测量值 x_t 对应的 FAU 应该是:

$$\mathrm{FAU} = \frac{1}{N} \sum_{i=1}^{N} y_i + \tilde{y} \tag{6}$$

我们采用极大似然法对上式中的未知参数进行求解,似然函数具有如下形式:

$$\ln(L(\boldsymbol{y}_D|\boldsymbol{x}_D,\theta)) = -\frac{1}{2}\ln|\boldsymbol{K}_N| - \frac{1}{2}\boldsymbol{y}_D^{\mathrm{T}}\boldsymbol{K}_N^{-1}\boldsymbol{y}_D \quad (7)$$

其中 θ 代表未知参数,比如径向基函数的参数和 δ_ω^2;L 表示 θ 的似然概率.

2.2 低维隐变量空间中的粒子滤波跟踪

跟踪变量是 $V=\{T,\mathrm{FAU}\}$,其中 T 是头部刚性运动参数,包括 x,y,z 三个方向的平移和旋转共 6 个参数.由于头部姿态 T 与表情动作单元 FAU 的变化相对独立,所以我们对它们采用两种不同的跟踪策略.对于 T 我们用梯度优化法进行计算,其目标优化公式为:

$$\arg\max_T \frac{1}{|A_0|}\| A(T,\mathrm{FAU}) - A_0 \|^2 \quad (8)$$

其中 $A(T,\mathrm{FAU})$是 FAU 的图像映射,由 Candide 模型[11]来计算.A_0 是人脸模板图像,可以从视频的第一帧手工标定获得;$|A_0|$是 A_0 中的像素数目.式(8)可以用一阶简化算法[12]求解.

对于 FAU 我们利用粒子滤波法在低维隐变量空间中求解.在 $t-1$ 时刻从 X 的取值空间里随机采样 J 个样本粒子$\{x_{t-1}^i \mid i=1,2,\cdots,J\}$,并采用随机行走模型来计算每个样本的状态迁移概率:

$$P(x_t^{(i)}|x_{t-1}^{(i)}) = N(x_{t-1}^{(i)},U) \quad (9)$$

其中 U 是随机行走方差,由实验确定.我们采用添加了置信度的似然概率来估计后验分布:

$$P(x_t^{(i)}|I_t) = \frac{1}{C}[\alpha P(I_t|x_t^{(i)}) + (1-\alpha)\tilde{\sigma}^2] \quad (10)$$

其中 C 是归一化常数.α 为置信度在后验分布中所占的比重,由实验确定.σ^2 是式(5)中的方差值.$P(I_t|x_t^{(i)})$为粒子 $x_t^{(i)}$ 对应的 FAU 图像映射与人脸模板的差值平方和.t 时刻 X 的取值为:

$$X_t = \sum_{i=1}^{J} P(x_t^{(i)}|I_t)x_t^{(i)} \quad (11)$$

综上所述,本文提出的 FAU 算法是一种结合了梯度优化和粒子滤波法的混合算法.它的主要步骤如表 1 所示:

表 1　混合跟踪算法

(1)在视频的第 1 帧对 X 和 T 进行手工初始化
(2)生成 N 个采样粒子$\{x^{(n)},\pi^{(n)},n=1,\cdots,N\}$.其中每一个粒子的 x 值设为 X 的初始化值,$\pi=1/N$
(3)对于视频的第 i 帧,$i>1$
　(a)使用梯度优化算法[12]计算当前帧的 T 值;
　(b)进行粒子的重采样:将 N 个粒子按 π 排序,将后 1/5 粒子用前 1/5 粒子替换;
　(c)进行粒子的采样:将每个粒子的 x 值按随机行走模型进行变迁(如公式 9);
　(d)更新粒子的权值 π:用式(6)计算每个粒子对应的 FAU 值,并结合 T 值计算该粒子的后验概率(如式(10)),并以后验概论的高低重新计算权值 π.把更新后的权值归一化,使$\sum\pi_i=1$;
　(e)根据式(11)计算本帧的跟踪结果:

$$E(x_i) = \sum\pi_i^{(n)} * x_i^{(n)}$$

3 实验

3.1 实验数据

我们用 130 万像素的 WebCam 在室内拍摄了两段人脸表情视频.视频中的人脸有连续的表情和头部姿态变化.视频图像的采集频率是 30 帧,分辨率为 320×240 像素,24 位真彩色.视频 1 的长度约 53 秒共 1605 帧,视频 2 的长度约 24 秒共 730 帧.我们从视频 1 中随机选取 108 帧图像,手工标定图像中人脸的 FAU 值.根据式(7)和这些标定的 FAU 值我们训练出 2 维隐变量高斯过程模型,其分布如图 3 所示.视频 2 被用来对跟踪算法进行测试.在手工标定 FAU 值时我们借助了上文提到的 Candide 模型.

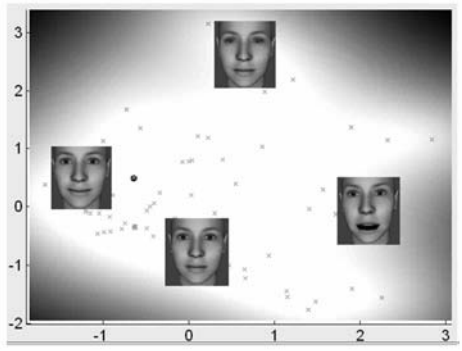


图 3　基于高斯过程的 2D 隐变量分布,其中亮度高的位置代表先验方差较小的 x

3.2 实验结果

我们分别采用梯度优化算法[12]*(简称 GD)、粒子滤波算法[2]**(简称 PF)和本文提出的算法对视频 2 进行 FAU 跟踪.我们选择眉毛降低、眉梢扬起、闭眼睛、嘴角拉伸、嘴角下沉和下颌降低这 6 种 FAU 作为跟踪对象,因为它们的组合成已经可以表达常规的人脸表情[2].我们从跟踪精度、跟踪平滑度两个方面来对结果进行分析.在跟踪精度方面我们采用 FAU 所覆盖的图像与人脸模板图像的差值平方和进行度量[2,12],见图 4,图 5;在平滑度方面我们采用跟踪结果的方差值和高频频域分量进行度量,见图 6,图 7.

图 4(a)显示了三种方法的跟踪误差曲线图,其中横轴代表图像帧的序号,纵轴代表跟踪误差.从该图可见,本文算法的误差曲线除在 70~130 帧高于上述两者算法之外,基本与粒子滤波法重合.而梯度优化算法的

* 我们分别实验了自变量扰动为 0.01, 0.03, 0.05, 0.07 的梯度优化算法,并从中选取跟踪误差最小的自变量扰动为 0.03 的梯度优化算法与本文算法比较

** 我们分别实验了粒子数为 200,重采样率为 0.2 或 0.3,采样方差为 0.05 或 0.0667 的粒子滤波算法,并从中选取了跟踪误差最小的重采样率为 0.2,采样方差为 0.05 的粒子滤波法来与本文算法比较

误差曲线则在大多数情况下高于其它两种算法.图 4(*b*)是一个箱线图,其中矩形中间的水平线分别代表三种算法的跟踪误差均值.总体上讲本文算法的精度与粒子滤波法基本持平,但明显高于梯度优化算法.然而本文算法具有图 4 不能反映的优点:即先验约束所带来的数值上难以体现的精度提升.图 5 可以显示这种优点.图 5(*a*)显示了视频 2 的第 663 帧,图像中的人脸有侧向偏转,但嘴部呈自然放松状,并无 FAU 变化.图 5(*b*)、(*c*)、(*d*)分别是根据梯度优化算法、粒子滤波算法和本文算法的跟踪结果所合成的去头部偏转的虚拟表情.可见,只有本文算法的跟踪结果才忠实体现了图 5(*a*)中的嘴部表情,而另外两种算法的跟踪结果都不约而同地将嘴部拉长了.但在图 4(*a*)中以竖直虚线标出的第 663 帧跟踪误差显示本文算法的误差反而高于粒子滤波算法.可见数值上较优的跟踪结果可能不如数值上较差的跟踪结果.

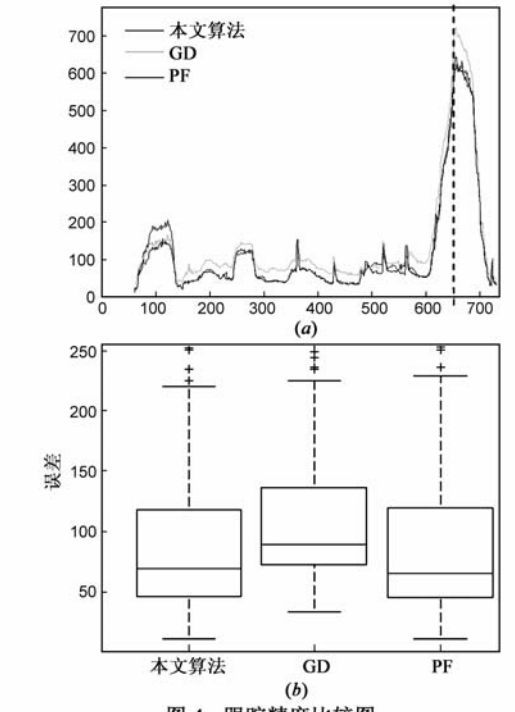


图 4 跟踪精度比较图

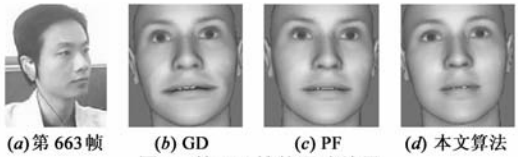


图 5 第 663 帧的跟踪结果

图 6 中有六个箱线图,分别对应 6 个 AU 的跟踪结果分布状态.其中每幅箱线图中的水平线代表跟踪结果的均值,矩形盒子反映了跟踪结果的方差,越长说明跟踪结果越分散,越扁说明跟踪结果越集中,而十字标记则代表离群点(outlier).从图中可见本文算法除闭眼睛 AU 之外,盒子的长度均明显低于另外两种算法.在精度较优的前提下,跟踪结果分布的相对集中反映了本文算法的跟踪平稳性.图 7 从频谱的角度对上述结论进行了验证.其中六个曲线图分别对应 6 个 AU 跟踪结果的 FFT 值.横轴代表频率值,纵轴代表分布比重.从图中可见除闭眼睛 AU 之外,本文算法跟踪结果中的高频变化分量明显低于另外两种算法.

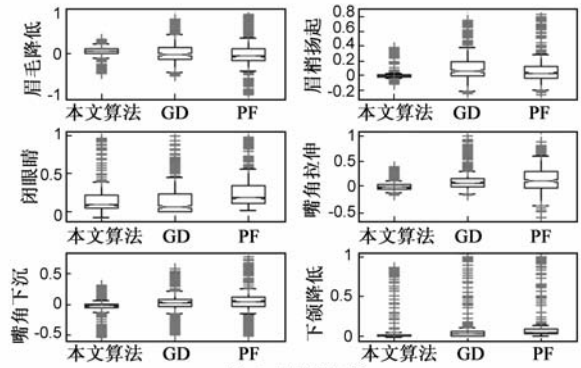


图 6 跟踪方差

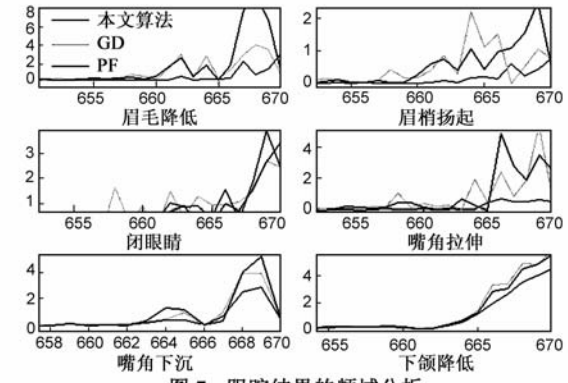


图 7 跟踪结果的频域分析

4 总结

针对表情动作单元跟踪中的“跟踪不平滑”和“变量缺乏约束”问题,本文提出了一种平滑且具有先验约束的 FAU 跟踪方法.它利用高斯过程的非线性降维特性来降低优化空间的维度,从而使跟踪过程更平滑;利用高斯隐变量空间的分布方差来对跟踪结果实施有效约束,从而降低跟踪过程中的非数值型误差;此外它还利用粒子滤波算法对非线性分布的采样能力来保证跟踪的精度.实验结果表明本文算法比传统的梯度优化和粒子滤波法具有更好的平滑性和跟踪精度.本方法可应用于表情识别和 3D 虚拟表情合成领域.

参考文献:

[1] Ekman P, Friesen W V. Facial Action Coding System: A Technique for the Measurement of Facial Movement[M]. Palo Alto, Calif, USA: Consulting Psychologists Press, 1978.

[2] Fadi Dornaika, Franck Davoine. Simultaneous facial action

tracking and expression recognition using a particle filter[A]. Proceedings of International Conference of Computer Vision (ICCV)[C]. Washington, DC, USA: IEEE Computer Society, 2005, 2: 1733 – 1738.

[3] Shaohua Kevin Zhou, Rama Chellappa, Baback Moghaddam. Visual tracking and recognition using appearance-adaptive models in particle filters[J]. IEEE Transactions on Image Processing, 2004, 13(11): 1491 – 1506.

[4] Neil D Lawrence. Gaussian process latent variable models for visualization of high dimensional data[A]. Advances in Neural Information Processing Systems 16[C], Cambridge, MA, USA: MIT Press, 2004. 329 – 336.

[5] David J C MacKay. Information Theory, Inference, and Learning Algorithms[M]. 3rd printing, Cambridge, UK: Cambridge University Press, 2003.

[6] Neil D Lawrence, Joaquin Quinonero-Candela. Local distance preservation in the GP-LVM through back constraints[A]. Proceedings of the 23rd International Conference on Machine Learning[C]. New York, NY, USA: ACM Press, 2006. 513 – 520.

[7] Tai-Peng Tian, Rui Li, Stan Sclaroff. Articulated pose estimation in a learned smooth space of feasible solutions[A]. Proc CVPR Learning Workshop[C]. Washington, DC, USA: IEEE Computer Society, 2005. 3. 50.

[8] Raquel Urtasun, David J Fleet, Aaron Hertzmann, Pascal Fua. Priors for people tracking from small training sets[A]. Proceedings of International Conference of Computer Vision (ICCV)[C]. Washington, DC, USA: IEEE Computer Society, 2005. 1. 403 – 410.

[9] Raquel Urtasun, David J Flee, Pascal Fua. 3D people tracking with gaussian process dynamical models[A]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. Washington, DC, USA: IEEE Computer Society, 2006. 1. 238 – 245.

[10] Martin D Buhmann, M J Ablowitz. Radial Basis Functions: Theory and Implementations[M]. Cambridge, UK: Cambridge University Press, 2003.

[11] M Rydfalk. CANDIDE, a parameterized face[R]. Report No. LiTH-ISY-I-866, Sweden: Dept of Electrical Engineering, Linköping University, 1987.

[12] T F Cootes, G J Edwards, C J Taylor. Active appearance models[J]. IEEE Trans, 2001, PAMI-23(6): 681 – 685.

作者简介:

王　磊　男,1978 年生于河南平顶山,博士研究生.研究方向为计算机图形图像处理.
E-mail: solehome@163.com

邹北骥　男,1961 年生于江西南昌,博士,教授,博士生导师.主要研究领域:计算机图形学、图像处理、软件工程技术等.
E-mail: bjzou@vip.163.com