

基于非线性 Boosting 回归的多视角人脸配准

苏延超¹ 艾海舟¹ 劳世²

摘要 人脸配准可以作为表情分析, 人脸识别等人脸相关研究的预处理步骤, 是人脸相关的计算机视觉研究的关键问题. 本文针对图像中水平视角在正负 45 度内的人脸配准问题, 利用基于 Haar 特征的非线性 Boosting 回归算法根据标定邻域内的局部纹理预测标定点的位移, 提出了一种新的基于经典活动形状模型 (ASM) 的实时多视角人脸配准算法. 在两个数据集上的测试实验表明, 该算法在速度、准确度和稳定性上都比经典的 ASM 算法有显著提高且优于近期的改进算法, 具有明显的实用价值.

关键词 人脸配准, 活动形状模型, Boosting 回归算法

中图分类号 TP391

Non-linear Boosting Regression for Multi-View Face Alignment

Su Yan-chao¹ Ai Hai-zhou¹ Lao Shi-hong²

Abstract Face alignment is a critical problem in many face related applications such as facial expression analysis, face recognition, etc. In this paper, we use local textures of each label point to predict the displacement of each label point by applying non-linear boosting regression based on Haar rectangle feature, and develop a novel real-time multi-view face alignment system based on Active Shape Model. Experiments on two independent datasets show that our algorithm is much faster, more accurate and robust than the classic Active Shape Model and outperforms recent improved algorithm too. It has significant practical value in real applications.

Key words face alignment, Active Shape Model, non-linear boosting regression

在人脸相关的计算机视觉研究中, 人脸配准通常是一个必要的前处理步骤. 例如三维人脸建模, 人脸姿态估计, 表情分析等应用都需要利用人脸配准提取人脸的形状信息. 现有的人脸配准算法大多基于点分布模型, 其中活动形状模型 (ASM)^[1] 和活动表观模型 (AAM)^[2] 是两个主流的研究方向.

活动形状模型和活动表观模型主要区别在于对图像纹理信息的利用. 活动形状模型利用标定周围的局部纹理模型搜索标定点的位置, 而活动表观模型则根据全局纹理的重构误差预测形状参数的变化. 活动表观模型对于形状变化比较鲁棒, 但对于光照较为敏感, 且容易限于现有数据集而推广性较差, 而活动形状模型配准时搜索的局部性决定其对形状初始化较为敏感.

针对活动形状模型研究者提出了许多改进, 其中^[3-4] 提出利用局部纹理分类器区分标定点的局部纹理和非标定点的局部纹理, 配准过程中在标定邻域内逐像素进行搜索. 由于利用区分性模型

代替原有的产生性模型, 局部纹理分类器能够更好的搜索到标定点的位置, 从而提高了算法的精度. 但精度提高的同时, 由于需要在特征点邻域内逐像素进行穷举搜索, 使得配准算法的计算复杂度大大增加. 此外, 由于采用局部搜索的策略, 算法的鲁棒性直接受搜索区域大小的影响, 而过大的搜索区域又会造成计算复杂度的增加.

在^[5]的工作中, 作者基于活动表观模型 (AAM), 提出利用非线性 Boosting 回归根据全局形状无关纹理预测形状参数变化量. 由于活动表观模型直接根据全局纹理预测形状参数的变化量而不是进行搜索, 因此不仅具有较低的计算复杂度, 而且不会受到搜索范围的影响. 受这一思想的启发, 我们认识到可以将活动形状模型 (ASM) 和非线性 Boosting 回归结合起来, 利用非线性 Boosting 回归函数根据标定点的局部纹理直接预测标定点的位移量, 从而得到标定点的位置. 这样采用回归方式避免了以往活动形状模型中在邻域内穷举搜索的过程, 从而大大降低了配准算法的复杂度, 避免了算法的鲁棒性受搜索区域大小的影响. 现有大多数人脸配准的研究都是针对正面人脸进行的. 对于多视角人脸配准, 人们给予的关注较少. 然而, 在一些实际应用如视频分析和检索中, 人脸并不能一直保持正面的姿态, 而大部分人脸的水平视角在正负 45 度之内. 多视角人脸配准面临的是人脸姿态变化引起的

收稿日期 XXXX-XX-XX 收修改稿日期 XXXX-XX-XX
Received Month Date, Year; in revised form Month Date, Year
国家自然科学基金 (60673107), 欧姆龙公司资助
Supported by National Natural Science Foundation of China
1. 清华大学计算机系 北京 100084 2. 日本欧姆龙公司核心技术研究所京都 619-0283 中国科学院自动化研究所模式识别国家重点实验室北京 100080
1. Science and Technology in Tsinghua University, Beijing 100084 2. Core Technology Center, Omron Corporation, Kyoto, JAPAN 619-0283
DOI: 10.3724/SP.J.1004.2008.xxxxx

二维形状的非线性变化和标定点处纹理的剧烈变化. 在以往的多数研究中, 人们将人脸的视角划分为若干个区域, 对每个视角区域中的人脸样本分别建模, 配准时选用相应视角的模型进行配准^[6-8]. 也有一些工作利用三维人脸模型辅助多视角人脸配准, 在配准过程中利用三维人脸模型恢复人脸的三维形状, 并进一步估计人脸姿态, 从而选取对应的活动形状模型进行配准^[9-10]. 也有研究直接利用三维人脸形状模型进行配准的^[11]. 针对多视角人脸二维形状的剧烈变化, 也有人^[12]利用分层形状进行配准. 文中首先根据器官的方向分类器寻找器官所在的方向并根据器官的纹理分类器的置信度确定器官所在的位置, 然后利用 ASM 方法在器官位置的约束下进行整体形状的配准.

在本文中, 我们针对一般图片中正负 45 度范围内的多视角人脸配准问题, 结合多活动形状模型配准与非线性 Boosting 回归的方法, 建立了一个准确、快速且鲁棒的多视角人脸配准算法. 在我们的方法中, 人脸的视角范围被划分为 3 个相互重叠的区间, 针对每个区间中的人脸样本学习出一套纹理回归函数, 对所有样本的形状统一建立形状模型, 配准时根据形状模型的参数估计人脸姿态并选取对应视角的局部纹理模型进行搜索.

1 全自动多视角人脸配准框架

利用活动形状模型进行人脸配准, 需要人脸形状和姿态的初始值. 这个初始化步骤由多视角人脸检测器完成. 人脸检测器检测出图像中人脸的大小和姿态, 初始化形状由形状模型的平均形状确定.

用活动形状模型进行人脸配准, 需要人脸形状和姿态的初始值. 这个初始化步骤由多视角人脸检测器完成. 人脸检测器检测出图像中人脸的大小和姿态, 初始化形状由形状模型的平均形状确定.

对于所有人脸样本, 我们将其标定点的坐标连接起来得到形状向量 $\mathbf{S} = (x_1, y_1, \dots, x_n, y_n)$ 并将所有人脸形状规范化到正切空间后建立 PCA 模型, 人脸的形状可以由姿态参数 q 和形状参数 p 确定:

$$\bar{\mathbf{S}} = T_q(U \cdot \mathbf{p} + \mathbf{S}) \quad (1)$$

其中 T_q 是以 q 为参数的正交变换, U 是形状主分量矩阵, $\bar{\mathbf{S}}$ 是平均形状.

形状模型将人脸形状的变化约束在 PCA 子空间中, 不仅降低了维度, 又可以确保人脸形状不会发生拓扑结构的变化.

图 1 给出了形状模型中不同参数下的三个形状. 由于视角变化过大, 简单的用一套局部纹理模型并不能覆盖局部纹理随视角的非线性变化, 因此我们将人脸的水平视角划分为 3 个相互交叠的区间, 如

表 1:



图 1 形状模型

Fig. 1 Shape Model

表 1 人脸视角划分

Table 1 View categories

视角	-1	0	1
角度范围	$-45^\circ \sim -15^\circ$	$-30^\circ \sim 30^\circ$	$15^\circ \sim 45^\circ$

每个视角区间 v 中, 对每个标定点 i 我们学习得到一个非线性回归函数, 根据该标定点周围的局部纹理预测该标定点的位移:

$$(\Delta x, \Delta y) = F_i^v(I_i) \quad (2)$$

2 局部纹理的 Boosting 回归

2.1 Boosting 回归函数

在活动形状模型配准中, 局部纹理模型用于在标定点邻域内搜索最符合局部纹理模型的标定点, 无论是产生式模型还是区分性模型, 为保证搜索的准确性, 我们必须在足够大的范围内进行搜索才能尽量避免落入局部最优, 然而搜索范围的增大会导致运算负担的增大, 因此搜索范围的设置需要权衡算法的准确性和运行效率. 而在活动形状模型配准框架中, 大部分运算用于局部搜索特征点的过程. 因此如何根据局部纹理找出标定点的位置是基于活动形状模型的配准算法改进的关键. 在本文中, 我们利用 Boosting 回归函数直接根据局部纹理预测标定点的位移替代搜索过程.

Boosting 回归函数是一系列非线性弱回归函数的和:

$$(\Delta x_i, \Delta y_i) = F_i(I_i) = \sum_k f_i^k(I) \quad (3)$$

其中的弱回归函数由 Boosting 算法从一个函数族中挑选出来. Boosting 学习过程是一个贪心算法, 每次从回归函数族中找出一个弱回归函数, 使得整个强回归函数的回归误差最小:

$$F_i^k(I_i) = F_i^{k-1}(I_i) + f_i^k(I_i) \quad (4)$$

其中:

$$f_i^k(I_i) = \arg \min \sum_i \|F_i^{k-1} + f_i^k(I_i) - (\Delta x_i, \Delta y_i)^T\|^2 \quad (5)$$

(5) 式等价于:

$$f_i^k(I_i) = \arg \min \sum_i \|f_i^k(I_i) - (\Delta x_i^k, \Delta y_i^k)^T\|^2 \quad (6)$$

其中:

$$\begin{aligned} (\Delta x_i^k, \Delta y_i^k)^T &= (\Delta x_i, \Delta y_i)^T - F_i^k(I_i) \\ &= (\Delta x_i^{k-1}, \Delta y_i^{k-1})^T - f_i^{k-1}(I_i) \end{aligned}$$

学习算法如算法 1:

算法 1 标定点处的回归函数的训练

Algorithm 1 Training the regression function

输入: 样本集 $\{(\Delta x_i, \Delta y_i, I)\}$ 和特征集 $\{f\}$

Start: $F_i = 0$

for $k = 0$ to N

$E^* = 0, f_i^{k*} = 0$

for $j = 1$ to N_f

根据样本学习基于第 j 个弱特征的弱

回归函数 $f_i^{k,j}$ 并计算相应的权重 $\omega_i^{k,j}$

计算该函数的回归误差 E

if $E < E^*$ Then

$f_i^{k*} = f_i^{k,j}, E^* = E$

end if

end for

$F_i = F_{i-1} + f_i^{k*}$

for $j = 1$ to N_s

$(\Delta x_i^{k+1}, \Delta y_i^{k+1}) = (\Delta x_i^k, \Delta y_i^k) - \omega_i^{k*} f_i^{k*}$

end for

end for

2.2 弱回归函数

利用标定点周围的纹理估计标定点的位移, 是一个从 $M \times M$ 维空间(局部纹理)到 2 维空间(标定点位移)的映射问题. 算法中的弱回归函数就是要根据样本学习这种映射关系. 由于标定点周围的局部纹理随标定点位移变化的非线性性, 直接利用线性回归等方法并不能很好的解决这个问题, 而直接利用任意非线性函数来拟合不仅会增加运算负担, 而且需要大量的样本进行训练, 并不可行. 因此受人脸检测 [13] 和配准 [4][8] 中相关工作的启发, 我们

选用 Haar 矩形特征对局部纹理进行映射, 将样本从 $M \times M$ 维的纹理空间映射到 1 维的特征空间, 然后通过训练集合优化函数映射学习得到弱回归函数:

$$f(I_i) = L(h(I_i)) \quad (7)$$

其中, $f()$ 是弱回归函数, $h()$ 是 Haar 矩形特征, $L()$ 是函数映射. 对于给定的 Haar 矩形特征, 一个弱回归函数的回归能力就取决于映射函数 $L()$, $L()$ 的选取要使对所有训练样本的回归误差最小:

$$L = \arg \max \sum_i \|(\Delta x_i^k, \Delta y_i^k)^T - L(h(I_i))\|^2 \quad (8)$$

通常情况下, L 的最优解都是非线性的, 为了降低回归函数的复杂度, 我们利用分段线性函数(查找表 LUT)来拟合非线性函数:

$$L(v) = \mathbf{l}_k, v \in [v_k, v_{k+1}) \quad (9)$$

(8) 式变为:

$$L = \arg \max \sum_k \sum_{h(I_i) \in [v_k, v_{k+1})} \|(\Delta x_i^k, \Delta y_i^k)^T - \mathbf{l}_k\|^2 \quad (10)$$

上式为关于 \mathbf{l}_k 的最小二乘问题, 解析解为:

$$\mathbf{l}_k = \frac{1}{n_k} \sum_{h(I_i) \in [v_k, v_{k+1})} \begin{bmatrix} x \\ y \end{bmatrix} \quad (11)$$

即 L 在每个区间内的值是所有特征值落在这个区间内的样本的位移量的期望值.

弱回归函数的结构如图 2:

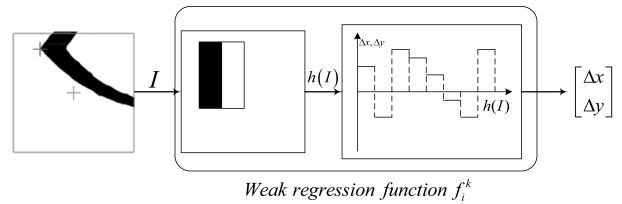


图 2 弱回归函数

Fig. 2 Weak regression function

2.3 基于非线性回归的活动形状模型配准算法

图 3 是多视角配准算法的流程:

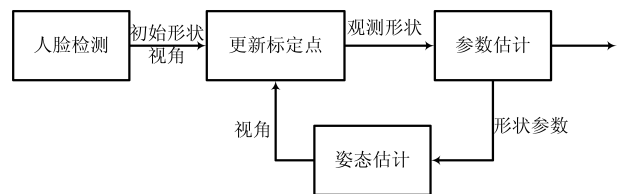


图 3 多视角配准流程

Fig. 3 Multi-view alignment flow

算法的初始形状和初始视角由多视角人脸检测程序提供 [14], 然后进入活动形状模型的迭代流程:

Step 1: 根据初始姿态, 对每个标定选取对应的局部纹理模型, 再根据标定点的局部纹理预测标定点的位移, 从而更新标定点的位置:

$$(x_i^*, y_i^*) = (x_i, y_i) + F_i(I_i) \quad (12)$$

Step 2: 根据形状模型估计人脸的形状参数和姿态参数, 并重构出人脸的形状

$$(p, q) = \arg \min \sum_i \|T_q(U_i \cdot p + \bar{S} - (x_i^*, y_i^*)^T)\|^2 \quad (13)$$

Step 3: 根据形状模型的参数估计人脸的姿态.

如此重复迭代直到人脸形状收敛.

在实验中, 我们直接根据形状参数 p 利用线性回归方法估计人脸的姿态, 然后根据估计出的姿态选取合适的局部纹理模型.

3 实验

3.1 实验数据

在实验中我们采用了两个数据集, 其中数据集 1 包含水平视角为 0° , 30° , 45° 的人脸图片共 3630 张, 图中人脸大小约为 200×200 , 其中视角的分布如表 2:

表 2 数据集 1 视角分布
Table 2 View categories of dataset 1

水平视角	0°	30°	45°
样本数量	1836	900	894

在每个视角中我们按比例选取 1/4 样本作为测试集合, 剩余 3/4 作为训练集合. 数据集 2 是选自人脸识别集合 Labeled Faces in the Wild^[15] 中的 600 张图像, 人脸图像大小约为 100×100 , 水平视角在正负 45° 范围内. 这个数据集中图像具有不同的光照、表情和复杂的背景, 而且分辨率较低, 较数据集 1 更难配准. 我们对每幅图像中的人脸手工标定了 88 个标定点. 数据集 2 中 3/4 用于训练, 剩余

1/4 用于测试. 在模型学习过程中, 每张图像每个标定点在手工标定的位置邻域内随机选取 75 个样本点作为训练数据, 每个标定点的回归函数包含 100 个弱回归函数.

3.2 对比试验

实验中, 我们将本文提出的算法和经典的 ASM 算法^[1] 以及基于局部纹理分类器及鲁棒参数估计的活动形状模型配准算法^[8] 在两个数据集上进行了比较. 选择算法^[8] 来做比较的原因是: 它是近年来在人脸配准方面有代表性的研究成果, 包含了算法^[4], 其中有系统的实验数据表明性能优于前期的有代表性的一些算法. 我们利用两个数据集对这三种算法分别进行了训练, 然后在测试集合上将各个算法的运行结果与手工标注结果进行了比较. 实验中, 如果配准算法迭代次数达到最大次数仍没有收敛, 或者收敛后配准结果与手工标定的结果的误差超过一定的阈值, 我们认为算法不收敛. 算法的收敛性能和时间对比如表 3 和表 4:

表 3 数据集 1 上性能对比

	收敛率	迭代次数	平均时间
经典 ASM 算法	75.2%	36	53ms
^[8] 中的算法	83.7%	13	403ms
本文算法	88.2%	23	65ms

表 4 数据集 2 上性能对比

	收敛率	迭代次数	平均时间
经典 ASM 算法	73.0%	39	54ms
^[8] 中的算法	84.3%	15	415ms
本文算法	89.1%	24	71ms

可见, 我们的算法收敛性能优于^[8] 中的算法且具有与经典 ASM 算法相当的速度.

对于算法收敛的样本, 我们统计了配准结果与手工标定数据误差的分布, 如图 4, 图 5.



图 6 数据集 2 中配准示例

Fig. 6 Results in dataset 2

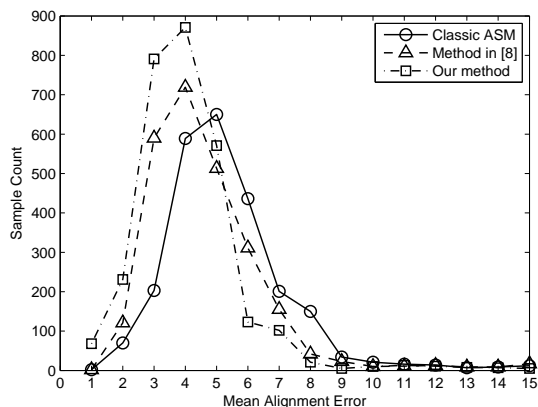


图4 数据集1上结果误差分布

Fig. 4 Error distribution on dataset 1

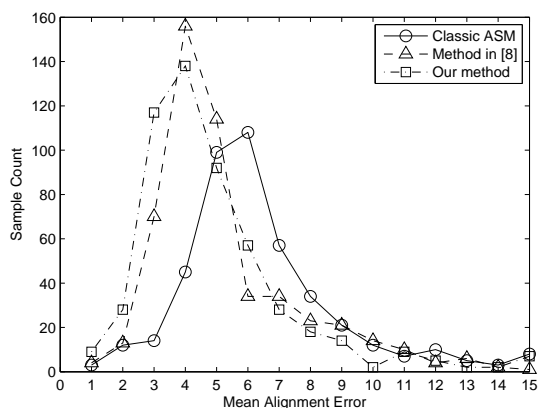


图5 数据集2上结果误差分布

Fig. 5 Error distribution on dataset 2

可见, 我们的算法在两个数据集中的性能都优于经典 ASM 算法^[1]和^[8]中的算法. 相比于最新的国际上其它相关工作, 由于采用的标定点的定义和数据集都不相同, 难以进行公允的系统性对比. 这里仅以文献^[12]中的实验结果为例说明: 该文选用了 Labeled Faces in the wild 中的一千多张人脸进行实验, 实验结果中平均点到点误差在 5 像素下的人脸占 74.7%. 而我们的算法是 73.4%, 与^[12]中的结果相当. 有关速度文中没有明确给出具体数据, 按照其描述可以推断配准时间应该不低于 200ms, 远比本文的算法慢. 图 6 是数据集 2 中的配准示例.

4 结论

本文针对水平视角在正负 45° 内的多视角人脸配准问题, 提出基于非线性 Boosting 回归的活动形状模型配准算法. 针对多视角人脸, 我们对近似正面 (正负 30°) 和侧面 ($-45^\circ \sim -30^\circ$, $30^\circ \sim 45^\circ$) 的人脸分别建立局部纹理模型, 而用一个统一的 PCA 模型表示人脸的形状. 配准过程中利用对应视角中基

于 Haar 矩形特征的非线性 Boosting 回归函数根据特征点周围的局部纹理预测标定点的位移从而更新标定点的位置, 然后利用形状模型估计形状参数, 并根据形状参数回归得到人脸的视角, 形成活动形状模型的迭代配准过程. 我们的算法较经典活动形状模型算法有很大的提高, 而且与近期改进的算法相比, 我们的算法在略微提高精度的同时大大提高了配准的速度, 具有明显的实用价值.

References

- 1 A Hill, T F Cootes, and C J Taylor. *Active shape models and the shape approximation problem*. In: **Proceedings of the 6th British Machine Vision Conference**. Birmingham, UK: Elsevier, 1995. 157–166.
- 2 T F Cootes. Statistical models of appearance for computer vision[Online]. <http://www.isbe.man.ac.uk/?bim/refs.html>, Sept.2001.
- 3 Yan Shui-chen, Li Ming-jing, Zhang Hong-jiang, Cheng Qian-sheng. *Ranking prior likelihood distributions for Bayesian shape localization framework*. In: **Proceedings of the 9th IEEE International Conference on Computer Vision**, Nice, France: IEEE, 2003. 51–53 vol. 1
- 4 Zhang Li, Ai Hai-zhou, Xin Sheng-jun, Huang Chang, Tsukiji Shuichiro, Lao Shi-hong. *Robust Face Alignment Based on Local Texture Classifiers*. In: **Proceedings of the 2005 IEEE International Conference on Image Processing**. Genoa, Italy: IEEE, 2005. 354–357, vol.2.
- 5 J Saragih, R Goecke. *A Nonlinear Discriminative Approach to AAM Fitting*. In: **Proceedings of the 11th IEEE International Conference on Computer Vision**, Rio de Janeiro, Brazil: IEEE, 2007. 1–8
- 6 Li Stan Z, Yan Shui-Cheng, Zhang Hong-jiang, Cheng Qian-sheng. *Multi-View Face Alignment Using Direct Appearance Models*. In: **Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition**. Washington, DC, USA: IEEE, 2002. 324–329.
- 7 Hu Chang-bo, Rogerio Feris, Matthew Turk. *Real-time View-based Face Alignment using Active Wavelet Networks*. In: **Proceedings of the 2003 IEEE International Workshop on Analysis and Modeling of Faces and Gestures**. Nice, France: IEEE, 2003. 215–221
- 8 Zhang Li, Ai Hai-zhou. *Multi-View Active Shape Model with Robust Parameter Estimation*. In: **Proceedings of the 18th International Conference on Pattern Recognition**. Hong Kong: IEEE, 2006. 469–468, Vol. 4.
- 9 Christian Vogler, Li Zhi-guo, Atul Kanaujia Siome Goldenstein, Dimitris Metaxas. *The Best of Both Worlds: Combining 3D Deformable Models with Active Shape Models*. In: **Proceedings of the 11th IEEE International Conference on Computer Vision**, Rio de Janeiro, Brazil: IEEE, 2007.1–7
- 10 Su Yan-chao, Ai Hai-zhou, Lao Shi-hong. *Multi-View Face Alignment Using 3D Shape Model for View Estimation*. In: **The 3rd IAPR/IEEE International Conference on Biometrics**. Alghero, Italy: Springer-Verlag, 2009.
- 11 Gu Lie and Takeo Kanade. *3D Alignment of Face in a Single Image*. In: **Proceedings of the 2006 IEEE International Conference on Computer Vision and Pattern Recognition**, New York: IEEE, NY, 2006. 1305–1312, vol.1

- 12 Liang Lin, Xiao Rong, Wen Fang Sun Jian. *Face Alignment Via Component-Based Discriminative Search*. In: **Proceedings of the 10th European Conference on Computer Vision**. Marseille, France: Springer-Verlag, 2008.72–85
- 13 P Viola and M Jones. *Rapid Object Detection using a Boosted Cascade of Simple Features*. In: **Proceedings of the 2001 IEEE International Conference on Computer Vision and Pattern Recognition**. Kauai Marriott, Hawaii, USA: IEEE.511–518, vol.1
- 14 Huang Chang, Ai Hai-zhou, Wu Bo, Lao Shi-hong. *Boosting Nested Cascade Detector for Multi-View Face Detection*. In: **Proceedings of the 17th International Conference on Pattern Recognition**. Cambridge, UK: IEEE. 2004. 415–418, vol.2
- 15 Huang Gary B, Manu Ramesh, Tamara Berg, Erik Learned-Miller. *Labeled faces in the wild: A database for studying face recognition in unconstrained environments*. University of Massachusetts, Amherst, Technical Report 07–49, October, 2007.



苏延超 清华大学计算机科学与技术系博士研究生. 2006 年获得清华大学学士学位. 主要从事模式识别, 人脸和人体的配准以及姿态估计工作.

E-mail:syc02@mails.tsinghua.edu.cn

(**Su Yan-chao** Ph.D. candidate in the department of computer science and technology. He received his bachelor degree from TsingHua University in 2006. His research interest covers pattern recognition, face and human alignment and pose estimation.)

degree from TsingHua University in 2006. His research interest covers pattern recognition, face and human alignment and pose estimation.)



艾海舟 《1985、1988、1991 年分别获得清华大学学士、硕士、博士学位, 1994.9-1996.8 在比利时布鲁塞尔自由大学作博士后研究, 现为清华大学计算机科学与技术系教授. 研究方向为计算机视觉. 本文通信作者.

E-mail:ahz@mail.tsinghua.edu.cn

(**AI Hai-Zhou** Received BS, MS, PhD in 1985, 1988, 1991 respectively all from Tsinghua University, as a Post Doctoral Researcher during 1994.9 1996.8 in Brussels University, Belgium, currently is a professor in Computer Science and Technology Department of Tsinghua University. Research area is computer vision.)



劳世竝 1984 年获得浙江大学学士学位, 1988 年获得日本京都大学硕士学位. 1992 年起在日本欧姆龙公司工作. 现为欧姆龙公司传感技术研究所高级研究员, 人脸图像处理的技术负责人.

E-mail:lao@ari.ncl.omron.co.jp

(**LAO Shihong** Received BS in 1984 from Zhejiang University, MS in 1988 from Kyoto University, Japan. Working in sensing technology laboratory of Omron corporation since 1988, currently he is a senior researcher, the project manager of face image processing.)