

一种改进的提取人脸面部特征点的 AAM 拟合算法

王 磊^{1,2}, 邹北骖^{1,2}, 彭小宁^{1,2}, 周 凌³

(1 中南大学信息科学与工程学院, 湖南长沙 410083 2 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; 3. 华为技术有限公司, 广东深圳 518129)

摘 要: AAM (Active Appearance Model) 是一种用来提取人脸特征点的有效方法, 由人脸动态表观建模和拟合算法两部分组成. 在多种 AAM 拟合算法中, 反向组合法以快速高效著称. 但在遇到外物遮挡时, AAM 算法的拟合效果会变差. 本文在反向组合法的基础上提出了一种基于分层细化掩模的改进算法. 实验结果表明, 该算法能较好地去除干扰并保留对拟合有用的信息, 具有较强的抗干扰鲁棒性.

关键词: AAM; 拟合算法; 反向组合法; 人脸特征点提取

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 0372-2112 (2006) 08-1424-04

An Improved AAM Fitting Algorithm for Extracting Human Facial Features

WANG Lei^{1,2}, ZOU Bei-jian^{1,2}, PENG Xiao-ning^{1,2}, ZHOU Ling³

(1. School of Information Science & Engineering, Central South University, Changsha, Hunan 410083, China; 2. State Key Lab of CAD & CG, Zhejiang University, Hangzhou, Zhejiang 310058, China; 3. Huawei Technologies Co., Ltd., Shenzhen, Guangdong 518129, China)

Abstract Active Appearance Model (AAM) is an efficient method for extracting human facial features. It includes active appearance models and the fitting algorithm. Within all kinds of fitting algorithms, the inverse compositional algorithm is one of the most efficient algorithms. However, the efficiency of the fitting algorithm will drop when some other objects occlude any parts of human face. An improved AAM fitting algorithm is presented. It not only keeps the superiority of the original inverse composition algorithm but also enhances the ability of anti-jamming. The experiments show that our algorithm can enhance the robustness of the AAM fitting algorithm and keep the useful information when fitting with occlusion.

Key words active appearance model (AAM); fitting algorithm; inverse compositional algorithm; extract facial feature

1 引言

面部特征点勾勒了人脸的主要形状特征, 一般位于人脸五官轮廓的尖端或突出处, 如内外眼角、嘴角和鼻尖等. 特征点的准确定位对于人脸识别、表情分析和基于视觉的人机交互等研究具有重要的意义. 现有的研究大多是采用无监督统计方法. 这些方法的主要思想是: 首先找出一组人脸基图像, 然后用基图像的线性组合来表示人脸. 近年来寻找快速有效的人脸特征点提取方法^[1-3]已成为计算机视觉研究的一个热点. AAM (Active Appearance Model)^[4-6]是一种快速有效的人脸特征点提取方法. 该方法由人脸动态表观建模和拟合算法两部分组成. 其基本思想是: 首先对一组已标记特征点的人脸图像进行训练, 根据

特征点位置和人脸表观来构造 AAM 模型 (人脸表观是人脸形状和纹理的组合; 将人脸特征点彼此相连组成三角形网络, 网格内部的像素光强就是人脸表观). 在拟合过程中, 通过调整模型参数使 AAM 的模型表观不断逼近输入图像, 从而实现输入图像中人脸特征点的自动定位和提取. 在此基础上可以进行更深入的分析, 比如人脸姿态的估计、人脸识别、表情分析、人脸图像编码等. AAM 算法最早由 F T Cootes 等人^[4]于 1998 年提出, 与其相近的概念有 Active Blobs^[7]和 Morphable Models^[8].

AAM 的拟合是一个非线性优化问题, 难以快速而有效地求解. 早期的 Levenberg-Marquardt 算法^[8]和 F. T. Cootes 等提出的随机梯度递减算法都是基于迭代优化, 在每次迭代中需要进行大量更新运算, 收敛速度不高. 卡耐基梅隆

大学的 Simon Baker 等人提出了反向组合算法^[6]. 它使 Hessian 矩阵和梯度可以提前计算, 有效减小了迭代计算量, 进而大大提高了拟合算法的速度. 但在人脸图像中存在外物遮挡, 上述算法的拟合效果都会显著下降. 针对外物遮挡和算法鲁棒性问题, 已有的算法^[9,10]都存在各自的不足. 文 [9] 的算法需要在 AAM 的训练过程中加入对遮挡的模拟, 因而大大提升了模型训练的难度; 文 [10] 的算法在利用掩模消除干扰的同时, 损失了大量对拟合有用的信息. 这些都不同程度地影响到了 AAM 拟合算法的效率.

本文在反向组合算法基础上, 结合掩模技术^[10]提出了一种基于分层细化掩模的 AAM 拟合算法. 该算法在解决遮挡问题的同时较好地保留了对拟合有用的信息, 能够在相对复杂和多变的环境中提取人脸特征点.

2 算法概述

反向组合算法^[6]由 Lucas-Kanade^[11]算法演化而来, 是一种快速有效的 AAM 拟合算法. Lucas-Kanade 算法通过求解式 (1) 的极小值实现输入图像与模板图像的最大匹配:

$$\sum_x [I(W(x; p)) - T(x)]^2 \quad (1)$$

其中 T 为模板图像, I 为输入图像. 反向组合算法不同于 Lucas-Kanade 算法的关键是: 不直接用 $p + \Delta p$ 来替代更新参数 p , 而是在 W 基础上进行更新 $W(x; p) \leftarrow W(x; p) \circ W(x; \Delta p)^{-1}$, W 是根据模板纹理的位置在输入图像对应位置上的纹理重采样. 反向组合算法将模板和输入图像的角色进行了调换, 将求式 (1) 的极小值转化为求式 (2) 的极小值:

$$\sum_x [T(W(x; \Delta p)) - I(W(x; p))]^2 \quad (2)$$

通过泰勒展开和求导得到变量 Δp 的表达式为:

$$\Delta p = H^{-1} \sum_x \left[\nabla I \frac{\partial W}{\partial p} \right]^T [T(x) - I(W(x; p))]$$

H 是 Hessian 矩阵: $H = \sum_x \left[\nabla I \frac{\partial W}{\partial p} \right]^T \left[\nabla I \frac{\partial W}{\partial p} \right]$. 这样处理使得 AAM 拟合计算中的 Hessian 矩阵和梯度值与位置变量 x 无关, 可以预先计算而无需迭代更新, 因此大大提高了算法的效率.

虽然反向组合算法的拟合速度比 Lucas-Kanade 算法快, 但它的适用条件仍然比较苛刻: 要求没有遮挡物, 光照均匀而恒定. 为了将该算法推向实际应用, 有必要进一步改进反向组合算法, 提高其鲁棒性.

3 算法改进

本文算法改进的基本思想是: 先计算输入图像的人脸表观与 AAM 平均人脸表观的差值图像, 得到两图像的不同部分. 再根据这些不同部分中心点的协方差, 并以各部分的区域面积为权重来计算不同部分的分散程度. 根据分散程度的不同采用不同大小的初始掩模. 在拟合过程中, 通过多层迭代判断和分块来进一步细化掩模, 使得掩模能根据遮挡出现的具体情况, 准确地设置在遮挡物体上, 最

终消除干扰.

在利用反向组合算法进行拟合计算的过程中, 若遇到外物遮挡, 模型与输入图像之间的误差值必然会大于正常情况下的误差值. 由于这类干扰出现时多集中在某个区域内, 因此考虑设置掩模 $M(x)$ 来消除干扰影响. 这样把求式 (2) 的极小值转化为求式 (3) 的极小值:

$$\sum_x M(x) \cdot [T(W(x; \Delta p)) - I(W(x; p))]^2 \quad (3)$$

对式 (3) 进行泰勒展开, 然后对 Δp 求导, 令求导式为零得到变量 Δp 的表达式:

$$\Delta p = H^{-1} \sum_x M(x) \cdot \left[\nabla T \frac{\partial W}{\partial p} \right]^T [I(W(x; p)) - T(x)]$$

其中 H 是 Hessian 矩阵:

$$H = \sum_x M(x) \cdot \left[\nabla T \frac{\partial W}{\partial p} \right]^T \left[\nabla T \frac{\partial W}{\partial p} \right] \quad (4)$$

根据干扰的区域性将图像分成 $B_1, B_2 \dots B_k$ 若干个子区域, 将整体掩模 $M(x)$ 分成区域掩模 M_i . 当遇到干扰集中的区域时掩模值 M_i 设置为 0 屏蔽该区域对拟合算法的影响; 其他区域的掩模值设为 1, 保留有效图像信息进行迭代计算. 如式 (5) 所示:

$$H = \sum_{i=1}^k M_i \sum_{x \in B_i} \left[\nabla T \frac{\partial W}{\partial p} \right]^T \left[\nabla T \frac{\partial W}{\partial p} \right] \quad (5)$$

分块掩模的引入提高了算法在有遮挡时的鲁棒性, 而且每个块区域 B_i 内的 $\nabla T \frac{\partial W}{\partial p}$ 值仍然能提前计算, 保留了原反向组合算法的速度优势. 文 [10] 的算法就采用了这种直观的掩模思想, 但它只有固定数目的单层掩模, 在抗干扰的同时也损失了大量有用信息.

本文在反向组合算法基础上, 结合掩模技术, 进一步改进了 AAM 拟合算法. 首先计算输入图像的人脸表观与 AAM 平均人脸表观的差值图像 $I(W(x; p)) - \bar{T}(x)$ ($\bar{T}(x)$ 是在 AAM 模型训练过程中得到的所有人脸表观的平均值), 得到两表观不同的部分 $Ea_1, Ea_2 \dots Ea_k$, 再根据式 (6) 计算不同部分的几何中心:

$$C_1 = (\bar{x}_1, \bar{y}_1), C_2 = (\bar{x}_2, \bar{y}_2), \dots C_k = (\bar{x}_k, \bar{y}_k) \\ \bar{x} = \frac{\int x dA}{A}, \bar{y} = \frac{\int y dA}{A} \quad (6)$$

在式 (6) 中, dA 是图像里的一个单位像素, A 是图像面积即像素点的个数. 积分是对所有像素坐标求和. 获取 C_k 的值是为了计算不同部分几何中心点的协方差, 如式 (7) 所示. S 的大小显示了不同部分的分离程度, 决定初始掩模 (即初始分块) 的大小.

$$S = \sqrt{\frac{\sum_{k=1}^n (\bar{x}_k - \bar{x})(\bar{y}_k - \bar{y})}{(k-1)}} \quad (7) \\ (\text{其中 } \bar{x} = \frac{\sum_{k=1}^n \bar{x}_k}{k}, \bar{y} = \frac{\sum_{k=1}^n \bar{y}_k}{k})$$

除了位置, 不同部分的面积同样影响着初始掩模的设置,

因此引入面积作为权重. 我们以不同部分的像素点个数 $A - E_{ak}$ 与总像素 $A - I$ 的比值作为权重 $W_k \left(W_k = \frac{A - E_{ak}}{A - I} \right)$, 再结合式 (7) 定义不同部分的分散度 D .

$$D = \sqrt{\frac{\sum_{k=1}^n W_k (\bar{x}_k - \bar{x})(\bar{y}_k - \bar{y})}{k-1}}$$

(其中 $\bar{x} = \frac{\sum_{k=1}^n \bar{x}_k}{k}, \bar{y} = \frac{\sum_{k=1}^n \bar{y}_k}{k}$)

根据不同的 D 值采用不同大小的初始掩模 MD_i . D 大, 采用小初始掩模, 则分块的块数多; D 小, 采用大初始掩模, 则分块的块数少. 在确定了初始掩模后, 再通过多层迭代判断和分块来进一步细化掩模. 先计算初始分块的各个子

区域 B_i 的误差值 E_i
$$E_i = \frac{\sum_{x \in B_i} [I(W(x; p)) - T(x)]^2}{\sum_{x \in B_i} [T(x)]^2 - \sum_{x \in B_i} [T(x)]^2}$$
, 分

别设上下限阈值. 若 E_i 小于下限值, 则说明误差值小, 设置该块掩模值 MD_i 为 1; 若 E_i 大于上限值, 则说明误差值大, 设置该块掩模值 MD_i 为 0. 在两阈值之间的区域, 需要进一步迭代判断和分块细化.

根据上述算法描述, 本文将 Hessian 矩阵的计算分为两部分:

$$H = H^* + \sum H^+ \quad (8)$$

其中 H^* 对应通过初次分块就能确定掩模值大小的区域. H^+ 对应需要迭代细分的区域. 其表达式为 (9):

$$H^* = \sum_{i=1}^{K_0} MD_i \sum_{x \in BD_i} \left[\nabla T \frac{\partial W}{\partial p} \right]^T \left[\nabla T \frac{\partial W}{\partial p} \right] \quad (9)$$

其中 K_0 是初始分块的块数, BD_i 指初始分块后的各个子区, MD_i 为初始掩模.

$$\text{令: } H_i = \sum_{x \in BD_i} \left[\nabla T \frac{\partial W}{\partial p} \right]^T \left[\nabla T \frac{\partial W}{\partial p} \right],$$

$$\text{得到: } H^* = \sum_{i=1}^{K_0} MD_i \cdot H_i.$$

继续推导 H^+ 的表达式: 将 E_i 值处于上限与下限之间的 BD_i 进一步分块, 假设块数为 K , 块区域为 B_{ij} , $BD_i = \sum_{j=1}^K B_{ij}$, 则由 (4) 可知 $H^+ = \sum_{x \in BD_i} M(x) \left[\nabla T \frac{\partial W}{\partial p} \right]^T \left[\nabla T \frac{\partial W}{\partial p} \right]$ 提出 $M(x)$ 得到:

$$H^+ = \sum_{j=1}^K M_j \sum_{x \in B_{ij}} \left[\nabla T \frac{\partial W}{\partial p} \right]^T \left[\nabla T \frac{\partial W}{\partial p} \right] \quad (10)$$

对式 (10), 令 $H_{ij} = \sum_{x \in B_{ij}} \left[\nabla T \frac{\partial W}{\partial p} \right]^T \left[\nabla T \frac{\partial W}{\partial p} \right]$, 得到: $H^+ = \sum_{j=1}^K M_j \cdot H_{ij}$. 将所有需要再分块区域的 H^+ 值累加, 得到 $\sum H^+$. 再根据式 (8) 就可得到 Hessian 矩阵的值. 由上述推

导可知, 因为 $\nabla T = \left(\frac{\partial T}{\partial x}, \frac{\partial T}{\partial y} \right)$ 是模板图像的梯度, $\frac{\partial W}{\partial p}$ 是在点

$W(x; 0)$ 即 x 点计算的 Jacobian 式, 所以 H_i 和 H_{ij} 仍然可以提前计算. 当掩模的细分迭代次数为 1 时, 本文提出的算

法退化为类似文 [10] 所描述的单层掩模算法; 当掩模的细分迭代次数大于 1 时, 对于掩模 M 被置为 0 的干扰分块, 由于 $\sum_{i=1}^K MD_i \geq M$, 所以可能丢失的非干扰信息比文 [10] 的算法少. 综上所述本算法保留了原反向组合算法的优越性; 同时由于掩模的迭代细分, 有效地保留了非干扰像素, 提高了抗干扰的精确度.

4 实验

4.1 实验方法

本文采用了标准 MM 人脸库 [12, 13] 作为 AAM 建模人脸训练集. 其中包括 7 位女性和 33 位男性的正面人脸图像和对应的特征点标记文件. 标记文件包含了对应人脸图像中 58 个面部特征点的位置. 这些特征点的分布如图 1 所示.

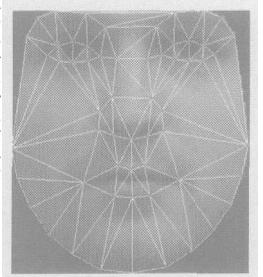


图 1

我们首先根据训练集中的人脸图像和标记文件进行 AAM 建模, 再分别采用反向组合算法、文 [10] 的算法和本文改进的 AAM 拟合算法对测试人脸图像集进行实验对比.

测试人脸图像集共有 240 幅正面人脸图像, 按照特征点被遮挡程度的不同可以分成三组: 无遮挡, 遮挡 10% 和遮挡 30%. 在完成对静态人脸图像的实验后, 我们还采集了有遮挡的视频序列进行实验, 视频序列为 320×240 像素, 15 帧/秒, 时间长度为 3 秒, 共 45 帧, 视频中的人脸从无遮挡转变到有遮挡, 再转变到无遮挡.

表 1 两算法对静态人脸特征点提取的统计结果

实验方法	无遮挡正面静态人脸图像提取正确率	遮挡 10% 特征点情况下的提取正确率	遮挡 30% 特征点情况下的提取正确率
原反向组合算法	95.83%	79.17%	58.33%
文 [10] 的算法	93.4%	86.06%	77.32%
本文算法	93.33%	88.33%	81.67%

4.2 实验结果

表 1 显示了反向组合算法、文 [10] 的算法和本文改进的 AAM 拟合算法对静态人脸特征点的提取结果, 其中无遮挡, 遮挡 10% 和遮挡 30% 特征点的人脸图像各 80 幅.

图 2 和图 3 分别显示了在有 10% 和 30% 特征点被遮挡的情况下, 三种不同算法的拟合结果图. (a) 为输入的静态图像, (b) 为采用反向组合算法得到的结果, (c) 为采用文 [10] 的算法得到的结果, (d) 为采用本文算法得到的结果, (e) 为本文算法产生的掩模图像.

图 4 显示了采用本文算法对人脸视频序列进行人脸特征点提取的结果.

5 结语

本文提出的改进算法保留了原 AAM 反向组合算法的

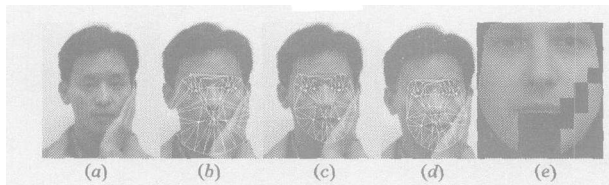


图 2 有 10% 特征点被遮挡的情况下的比较结果

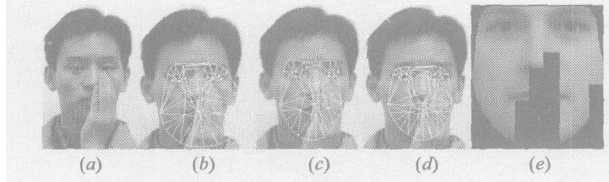


图 3 有 30% 特征点被遮挡的情况下的比较结果



图 4 对人脸视频序列进行人脸特征点提取的结果

优点,能避免在迭代过程中重复计算 Hessian 矩阵.本文提出的分层细化掩模法在去除遮挡干扰的同时,有效地保留了有效图像信息.实验表明它比原始的 AAM 反向组合算法和文 [10] 的算法更鲁棒.随着特征点被遮挡程度的增大,它的优势更加明显.但是该算法也存在不足,比如对因头部姿态变化而引起的自遮挡情况特征点提取效果不佳.下一步将围绕自遮挡问题进一步改进算法.

参考文献:

- [1] 宋海娜,匡纲要,等.基于对称变换与高斯微分的人脸定位新方法[J].电子学报,2003,31(9):1433-1436
Song Haina, Kuang Ganyaq, et al. A new face localization method based on symmetry transformation and Gaussian-derivatives[J]. *Acta Electronica Sinica*, 2003, 31(9): 1433-1436 (in Chinese)
- [2] 彭进业,俞卞章,等.多尺度对称变换及其应用于定位人脸特征点[J].电子学报,2002,30(3):363-366
Peng Jinye, Yu Bianzhang, et al. Multiscale symmetry transform with application to location of feature points on human face image[J]. *Acta Electronica Sinica*, 2002, 30(3): 363-366 (in Chinese)
- [3] 甘俊英,张有为,等.自适应主元提取算法及其在人脸图像特征提取中的应用[J].电子学报,2002,30(7):1013-1016
Gan Junying, Zhang Youwei, et al. Adaptive principal components extraction algorithm and its application in the feature extraction of human face[J]. *Acta Electronica Sinica*, 2002, 30(7): 1013-1016 (in Chinese)
- [4] G J Edwards, C J Taylor, T F Cootes. Interpreting face images using active appearance models[A]. 3rd IEEE International Conference on Automatic Face and Gesture Recognition Proceedings[C]. Washington, DC, USA: IEEE Com-

puter Society, 1998: 300-305

- [5] T F Cootes, G J Edwards, C J Taylor. Active appearance models[J]. *IEEE Trans*, 2001, PAM 1-23(6): 681-685
- [6] IM Athwary, S Baker. Active appearance models revisited[J]. *International Journal of Computer Vision*, 2004, 60(2): 135-164
- [7] S Sclaroff, J Isidoro. Active blobs[A]. 6th International Conference on Computer Vision Processings[C]. New York, NY, USA: Elsevier Science Inc, 1998: 1146-1153
- [8] V Blanz, T Vetter. A morphable model for the synthesis of 3D faces[A]. *Proceeding of SIGGRAPH Proceedings[C]*. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co, 1999: 187-194
- [9] R Gross, IM Athwary, S Baker. Constructing and fitting active appearance models with occlusion[A]. *Conference on Computer Vision and Pattern Recognition Workshop Proceedings[C]*. Washington, DC, USA: IEEE Computer Society, 2004: 72-72
- [10] Takahiro Ishikawa, Iain M Athwary, Simon Baker. Efficient Image Alignment with Outlier Rejection[R]. Pittsburgh Carnegie Mellon University Robotics Institute, 2002
- [11] S Baker, IM Athwary, Lucas Kanade. 20 years on: A unifying framework[J]. *International Journal of Computer Vision*, 2004, 56(3): 221-255
- [12] M B Stegmann, B K Ersbøll, R Larsen. FAME - a flexible appearance modeling environment[J]. *IEEE Trans*, 2003, *Medical* 22(10): 1319-1331
- [13] M M Nordstrom, M Larsen, J Sierakowski, M B Stegmann. The MM Face Database[DB/OL]. <http://www.imm.dtu.dk/~aam/>, 2004-05

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: sokehon@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