

基于双目被动立体视觉的三维人脸重构与识别

周佳立¹ 张树有¹ 杨国平²

摘要 提出一种基于双目被动视觉的三维人脸识别方法, 该方法采用非接触式的人脸信息采集技术, 利用图像中弱特征检测实现双目视觉中的人脸检测与初步视差估计, 运用基于复小波的相位相关技术对人脸表面进行亚像素级小区域匹配, 重建人脸三维点云信息. 通过可训练次数的神经网络技术实现多层次人脸曲面重建, 并结合人脸 2D 图像对重构曲面进行仿射归一, 继而迭代地进行特征提取与识别过程. 实验结果表明, 双目视觉方法使人脸信息采集过程友好隐蔽; 在对应点匹配中, 运用复小波的相位相关算法可获得密集的亚像素精度配准点对, 用神经网络方法可正确重建人脸曲面. 识别过程对环境以及人脸位姿表情等鲁棒性强. 该系统成本十分低廉, 适合在许多领域推广应用.

关键词 双目视觉, 三维人脸, 重构, 识别, 神经网络

中图分类号 TP391

A 3D Face Reconstruction and Recognition Method Based on Passive Binocular Stereo Vision

ZHOU Jia-Li¹ ZHANG Shu-You¹ YANG Guo-Ping²

Abstract In this paper, we present a 3D face recognition method based on passive binocular stereo vision. We introduce non-contact face information-collecting technique and use weak feature-detection method on static image to achieve face disparity estimation. Furthermore, we employ phase-correlation in complex wavelet to perform sub-pixel small region matching on the surface of human face, and reconstruct 3D point cloud information of human face. Neural network is applied to obtain multilevel face surface reconstruction. Combining with 2D image of human face, we affine and normalize the reconstructed surfaces, carry out feature extraction and recognition. The experiment shows that this method is robust against surroundings, as well as the position and expression of human face. Moreover, it has high accuracy and quick recognition speed.

Key words Binocular stereo vision, 3D face, reconstruction, recognition, neural network

人脸识别作为生物特征识别的一个重要方面, 在安全验证、信用卡验证、身份识别、银行和海关的监控、档案管理、人机交互等领域具有广阔的应用前景. 与指纹识别、虹膜识别、DNA 识别等技术相比, 人脸图像来源广泛, 并且使用非接触式的采集方式, 整个识别过程具有更加隐蔽、友好和便捷的特点, 更易为用户所接受^[1].

然而, 目前的人脸识别主要还是针对二维图像或二维动态视频序列进行研究. 二维图像识别技术在其他领域已有很多的应用, 但是由于人脸是个塑

变体, 使得通过二维图像识别技术对人脸识别存在困难. 另外, 基于二维图像的人脸识别, 不可避免受到环境光线、背景、视角等和人脸的姿态、表情、遮挡等不利影响, 因而其识别精度很难有进一步的提高^[2].

为了克服基于二维图像人脸识别技术的不足, 有研究者已经开始对三维人脸识别技术进行研究. 其中, Bronstein 提出了一个 3D 人脸识别方法^[3], 该方法允许与面部表情有关的变形, 把 3D 人脸数据转换成“特征形式”, 使它成为模型的形状变化的不变量. Gokberk 把基于扩展的高斯图像、ICP 匹配、距离轮廓、PCA 和线性判别式分析 (LDA) 的方法分别用作 3D 人脸识别^[4], 并使用 Beumier 和 Acheroy 使用的数据. Lee 提出了基于人脸 8 个特征点的曲率值的 3D 人脸识别方法^[5]. Lu 和 Jain 发展了以前基于 ICP 的识别方法, 用它清楚地处理人脸面部表情的变化^[6]. Russ 提出用 Hausdorff 距离匹配 3D 人脸数据的距离图像表示^[7]. 用类似于 ICP 中的迭代记录程序使被检验者的数据列适应图库中的数据列. Pan 用特征脸与 3D 数据深度图像的映射进行匹配^[8]. 把鼻尖作为中心点和数据列的对

收稿日期 2008-04-03 收修稿日期 2008-08-24

Received April 3, 2008; in revised form August 24, 2008

国家自然科学基金 (50775201), 国家科技支撑计划 (2006BAF01A45-09), 浙江省科技支撑计划 (2008C21084), 浙江省自然科学基金 (Z107416) 资助

Supported by National Natural Science Foundation of China (50775201), National Science & Technology Pillar Program (2006BAF01A45-09), Science & Technology Pillar Program of Zhejiang Province (2008C21084), and Natural Science Foundation of Zhejiang Province (Z107416)

1. 浙江大学 CAD & CG 国家重点实验室 杭州 310027 2. 浙江中医药大学信息技术学院 杭州 310053

1. State Key Laboratory of CAD & CG, Zhejiang University, Hangzhou 310027 2. College of Information Technology, Zhejiang Chinese Medical University, Hangzhou 310053

DOI: 10.3724/SP.J.1004.2009.00123

称轴, 将人脸数据映射成圆形的距离图像. Chang 将多区域方法引入到 3D 人脸识别中^[9]. 它是一种分类器集合体方法, 对鼻子周围的多个重叠子区域分别用 ICP 进行匹配, 并合并多重 3D 匹配的结果. Passalis 采用的是带注释的可变模型研究 3D 人脸识别, 在统计学的基础上对训练集计算出平均 3D 人脸^[10].

对三维人脸识别技术而言, 存在的最大困难是人脸三维点云数据的获得. 近几年光学三维形貌测量技术有了很大的进展, 已发展出一些具有实用潜力的三维形貌测量技术, 如结构光投影相移技术、结构光投影傅里叶变换技术等, 使得三维人脸的识别成为可能^[11]. 但是, 通过向人脸投射结构光事实上已经是一种接触式的识别方法, 对被识别者不够友好, 并且也失去了信息隐蔽采集的能力, 这样在很大程度上减弱了较之指纹识别等的优势. 因此, 如何快速高效并且隐蔽地获取被识别者的三维人脸信息引起了人们的重视.

目前, 越来越多的人将目光转向基于图像恢复人脸几何的方法来获得非接触式三维人脸信息, 其中一类方法要求输入的图像具有预定义结构^[12], 如采用正交视图, 从图像中提取出轮廓, 用于修改一般人脸模型, 得到最终结果. 另一类方法是先精确重建出人脸上的少数三维特征点, 然后用变分技术修改一般人脸模型^[13]. 然而, 这样的方法或者需要被识别者充分地配合, 或者只能准确获得少数的几个特征点的三维信息, 难以达到自动建模的效果.

在识别方面, 采集的人脸三维点云数据量通常十分庞大, 不宜直接进行储存和识别比对, 因此, 人们考虑通过曲面拟合的方式提取点云的整体信息进行比对与储存. 然而, 这些方法构造曲面随意度大, 对其归一化困难, 并且由于点云获取的随意性, 使得难以提出行之有效的尺度概念来对不同人脸进行识别区分. 因此, 如何通过采集的三维点云信息, 提取出易于储存与识别比对的人脸特征是亟待解决的重要问题.

本文介绍一种基于双目视觉的三维人脸识别技术, 通过运用训练完成的人脸分类器对视频流进行基于弱特征的人脸检测, 确定双目视觉中人脸的大致视差, 利用新的无接触式双目立体视觉技术对人脸三维点云数据进行快速密集采集, 提出人脸三维点云数据基于神经网络的曲面拟合技术, 结合 2D 图像信息, 对曲面进行仿射归一, 提取特征参数进行比对识别, 解决了传统人脸识别系统对环境以及人脸姿态表情等鲁棒性差的问题, 获得较高的识别正确率.

1 运用双目视觉系统获取人脸图像信息

采用上下两平行光轴摄像机建立双目立体视觉系统, 如图 1 所示. 由于在光轴距离小的情况下, 可以近似地将上下采集图中对应点为中心的小区域视为相互的一个平移, 因此, 希望它们的光轴距离 (基线距离) 越短越好. 在系统中, 基线距离即为摄像机的宽度, 取 60 mm. 这也正好是人的双眼的大致距离, 符合仿生学原理. 基于这样的假定, 可通过文章将要介绍的复小波相位相关算法自动寻找上下图中每个点的共轭对, 实现图像相应点亚像素精确匹配.

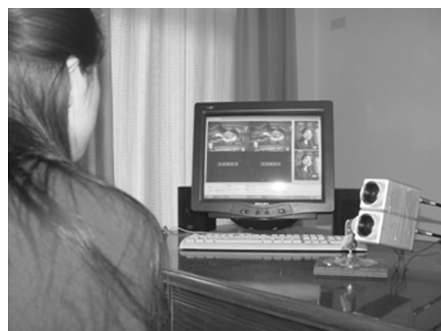


图 1 双目立体视觉系统

Fig. 1 Binocular stereo vision system

对上下摄像机进行立体视觉标定^[14-15], 由于受到镜头畸变的影响, 并且在实际应用中双目光轴往往难以达到精确平行要求, 因此外极线通常不垂直于图像 X 轴. 通过矫正图像对 (如图 2 (见下页) 所示) 可解决这一问题. 因而, 在后继的共轭点匹配中, 只需对 Y 轴附近进行扫描即可.

对于标定矫正后的双目视觉系统, 当上下图中的共轭点匹配精度达到 0.1 像素时, 利用双目立体视觉原理可精确获得空间点的三维深度信息, 误差不超过 0.5 mm. 所以, 双目视觉系统的关键技术在于图像共轭点的精确匹配. 需要指出的是, 视觉系统采样的正确匹配需要物体表面有一定量的纹理信息, 否则相位相关算法将失败. 由于人脸表面蕴涵着丰富的纹理信息, 因此可较好地满足视觉采样要求. 在很少的情况下, 如皮肤细腻光洁的年轻女性, 匹配过程遇到困难. 后文将介绍说明人脸识别过程只受到了不多的影响.

在共轭点匹配前, 我们需要了解人脸在图像中的大致区域与对应视差. 这样, 匹配的搜索深度与范围将大大减少, 既提高了效率又减少了匹配时的误判与拒判率. 文章运用 Lienhart^[16-17] 介绍的基于灰度图像中弱特征检测方法对视频流进行实时人脸检测, 对于训练完成的多角度人脸 AdaBoost 分类器, 单帧图像人脸检测准确率可达 90%. 当人脸图像进入视觉系统中时, 系统对每一帧图像进行

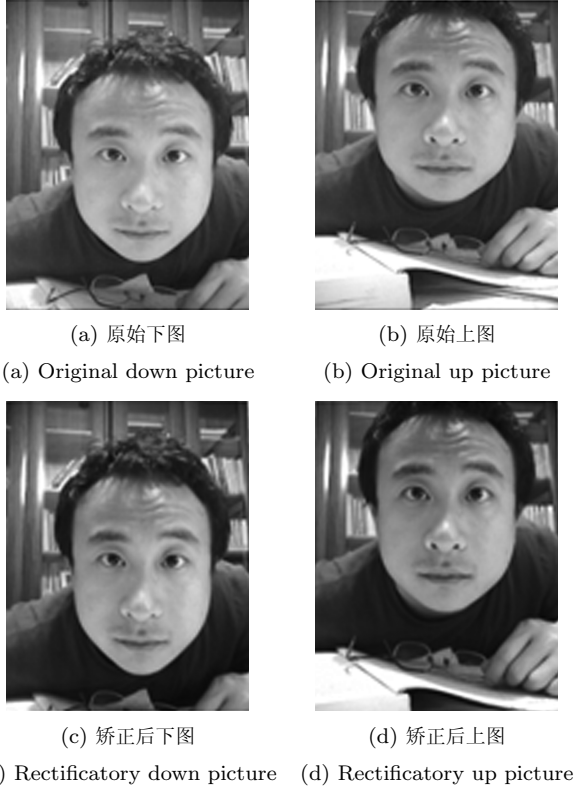


图 2 矫正图像对

Fig. 2 Pairs of rectificatory images

检测, 从检测结果中挑选最符合人脸逻辑的图像作为视觉系统的采样, 为后继的三维重建提供样本. 这里所谓的符合逻辑是指, 上下视频流中检测到的人脸区域的大小与其中中心 X 轴坐标需基本一致, 视差在一定的范围并且人脸区域需整个落在视频内. 此时记录人脸检测结果并估计人脸视差. 如图 3 所示, 矩形区域为自动检测到的人脸区域.

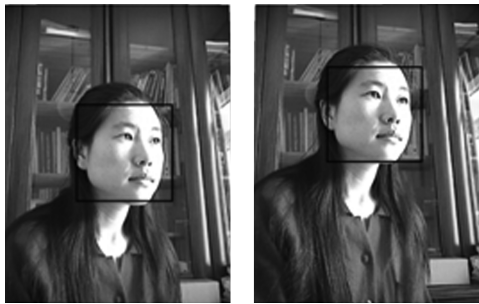


图 3 人脸检测与视差估计

Fig. 3 Face detection and disparity estimation

2 三维人脸点云信息重建

2.1 基于复小波变换的相位相关算法

给定两幅大小均为 $(2M+1) \times (2N+1)$ 的灰度图 $f_1(m, n)$ 、 $f_2(m, n)$, 定义它们在连续域中的对

应函数仍为 $f_1(x, y)$ 、 $f_2(x, y)$, 为方便叙述文章考虑连续函数的情况, 假设 f_2 为 f_1 的平移, 平移量为 (δ_1, δ_2) , 则

$$f_2(m, n) = f_1(m - \delta_1, n - \delta_2) \quad (1)$$

设 ϕ 为二维复小波基, 常用的小波基 $\phi(x, y)$ 多为变量可分离的二元函数: $\phi(x, y) = \phi_1(x)\phi_2(y)$, 在实际应用时采用符合人眼识别规律的 Morlet 复小波

$$\phi_1(t, \omega) = \phi_2(t, \omega) = e^{-\frac{t^2}{2}} e^{i\omega t} \quad (2)$$

对于连续域中的函数 f_1 、 f_2 , 在 $L^2(\mathbf{R})$ 上进行积分小波变换解 (Integral wavelet transform, IWT)

$$f(m, n) \rightarrow (W_\phi f)(b_1, b_2, a) \quad (3)$$

根据小波变换的平移性可知

$$(W_\phi f_2)(b_1, b_2, a) = (W_\phi f_1)(b_1 - \delta_1, b_2 - \delta_2, a) \quad (4)$$

当 $\delta_1 \ll a$, $\delta_2 \ll a$ 时,

$$\begin{aligned} (W_\phi f_2)(b_1, b_2, a, \omega_1, \omega_2) &\approx \\ (W_\phi f_1)(b_1, b_2, a, \omega_1, \omega_2) &\cdot e^{-i\omega_1(\frac{\delta_1}{a}) - i\omega_2(\frac{\delta_2}{a})} \end{aligned} \quad (5)$$

在实际匹配过程中, 采用从粗到细的搜索策略, 使得相对于给定的分辨率 a 、 δ_1 、 δ_2 总是很小, 式 (5) 的估计是合适的. 计算 f_1 、 f_2 的交叉能量谱 $\mathbf{C}$ 并归一化后记为

$$\hat{\mathbf{C}}(\omega_1, \omega_2) = \frac{\mathbf{C}(\omega_1, \omega_2)}{|\mathbf{C}(\omega_1, \omega_2)|} \approx e^{i\omega_1(\frac{\delta_1}{a})} \cdot e^{i\omega_2(\frac{\delta_2}{a})} \quad (6)$$

对于一定范围的 ω_1 、 ω_2 , 进行 $2D$ 离散, 记 $\omega_1 = (2\pi/N_1)k_1$, $\omega_2 = (2\pi/N_2)k_2$, 其中 $N_1 = 2M+1$, $N_2 = 2N+1$, $-M \leq k_1 \leq M$, $-N \leq k_2 \leq N$, 傅里叶反变换得

$$\begin{aligned} \hat{\mathbf{c}}(n_1, n_2) &= \frac{1}{N_1 N_2} \sum_{k_1 k_2} e^{i\frac{2\pi}{N_1} k_1 \frac{\delta_1}{a}} \times \\ &e^{i\frac{2\pi}{N_2} k_2 \frac{\delta_2}{a}} \cdot e^{i\frac{2\pi}{N_1} k_1 n_1} \cdot e^{i\frac{2\pi}{N_2} k_2 n_2} \end{aligned} \quad (7)$$

理论上, $\hat{\mathbf{c}}(n_1, n_2)$ 的虚部应等于 0, 然而实际应用中, 通常只取其部来用二维 Dirichlet 函数拟合

$$\begin{aligned} \text{Re}[\hat{\mathbf{c}}(n_1, n_2)] &\approx \\ \sigma &\cdot \frac{\sin\left[\pi\left(n_1 + \frac{\delta_1}{a}\right)\right]}{\sin\left[\frac{2\pi}{N_1}\left(n_1 + \frac{\delta_1}{a}\right)\right]} \cdot \frac{\sin\left[\pi\left(n_2 + \frac{\delta_2}{a}\right)\right]}{\sin\left[\frac{2\pi}{N_2}\left(n_2 + \frac{\delta_2}{a}\right)\right]} \end{aligned} \quad (8)$$

在理想情况, $\sigma = 1$. 在实际应用中, 由于存在噪声等的影响, σ 通常小于 1. 运用 Levenberg-Marquardt 算法对 $\text{Re}[\hat{\mathbf{c}}(n_1, n_2)]$ 中心 5×5 的整数点数值进行非线性最小二乘拟合, 得到包括 σ , δ_1 , δ_2 的参数估计.

2.2 人脸图像共轭点对匹配过程与重建

对人脸图像进行 4 层复小波分解, 在像素匹配阶段采用简化的算法: 从最底层开始, 只比对 $\text{Re}[\hat{c}(n_1, n_2)]$ 中心附近的点, 找出最大值, 而由该最大值作为 σ 的近似, 其位置即为相应的视差估计, 放大后代入上一层作为新的初始估计, 进行新的计算, 直到最顶层为止. 因此, 在进行亚像素匹配时, 视差估计的误差应该小于 1 个像素. 继而通过几次非线性最小二乘拟合迭代过程, 获得最终的亚像素级视差估计. 这样的迭代过程进行两次就可趋于收敛.

根据以上的匹配方法, 对于大部分点都可自动获得正确的视差估计, 然而由于遮挡、噪声以及缺少纹理等情况的影响, 在由粗到细的金字塔匹配过程中, 对于某些点对将出现错误估计, 进而导致亚像素匹配无法收敛等情况的发生. 不仅降低了自动配准的鲁棒性, 而且在很大程度上减弱了三维重建的可靠性和精准度.

为克服以上问题, 引入顺序匹配、连续性、相关性以及外极线等约束条件, 作为对匹配关系的可信度评估. 所谓顺序匹配约束是指, 对于上图中的两点, 其下图中的匹配点应保持原来的顺序; 连续性是指, 对于上图中接近的两点, 其相应视差也应该相似; 相关性约束即可用 σ 值来表示, 当 σ 值越大, 说明这两点的匹配相关性越大; 而极线约束是指匹配点将拥有基本相似的横坐标.

因此, 整个匹配过程可改进为以下算法:

步骤 1. 对于每一层, 匹配并估计每一未标记点的 σ 值. 如果 σ 小于某一阈值, 则对该点进行标记并转入步骤 2. 否则, 考虑其顺序约束, 取与其同列并已匹配的最近未标记点, 如果顺序匹配错误, 则进行标记并转入步骤 2. 否则, 考虑其连续约束情况, 取其附近 5×5 以内所有已匹配而未标记点, 取其视差平均, 比较与该点差的绝对值, 如大于某一阈值, 则对该点标记, 转入步骤 2. 最后, 如果匹配点偏离外极线过大, 则标记该点并转入步骤 2.

步骤 2. 对于标记的点, 考虑其附近 5×5 以内所有未标记点, 取其关于 σ 值的相对加权平均作为该点的初始估计, 然后进行亚像素匹配. 对其结果进行相应的约束检测. 如仍被标记, 则将初始估计作为其相应视差.

对图 4(a) 所示上下视觉采样进行人脸区域的点集匹配, 并根据视觉原理重建三维点云信息. 结果如图 4(b) 所示. 以上方法由于采用了顺序匹配、连续性、相关性以及外极线等约束条件, 因此, 对于已标记的匹配点将获得很好的精度要求, 通常误差被控制在 0.1 像素以内. 而由于采用高匹配相关性的点对作为种子, 并且采用由粗到细的匹配策略, 使得最终正确匹配率可达到 95% 以上, 只有很少的一部

分会出现误匹配情况, 即作为最终未标记点而被自动识别出来, 这些点可通过插值等方法给出近似的结果.

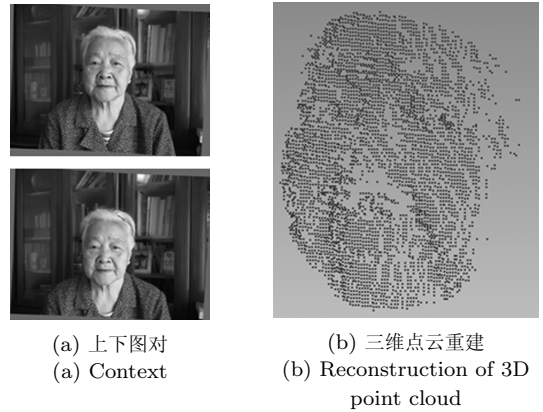


图 4 人脸三维点云重建结果

Fig. 4 3D point cloud reconstruction of the face

3 三维人脸曲面信息重建与识别

3.1 基于神经网络的人脸曲面重建

将上节获取的三维点云信息的 x, y 坐标作为神经网络的输入, 而将 z 坐标作为输出, 构成二输入、一输出的网络系统, 网络的输入与输出形成映射关系, 设这样的映射关系为一有限支单值连续函数 $l = \hat{f}(\mathbf{X})$, $\mathbf{X} = (x, y)$.

事实上, 对于两层单输出 MLFN (Multilayer feedforward neural network), 无论隐层取 Sigmoid 函数、小波函数还是 RBF (输出层皆取线性函数), 只要隐层神经元足够多, 就能以任意精度逼近任何有限支单值连续函数.

但是, 由于函数 $l = \hat{f}(\mathbf{X})$ 不是非常平滑, 如采用局部接收场的 RBF 或小波网络, 网络规模 (即隐层神经元数) 随输入向量维数的增加而指数增加. 为控制规模, 本文采用全局接收场的 Sigmoid 网络.

设计两层单输出且隐层取 Sigmoid 函数、输出取线性函数的 MLFN, 其映射函数可表示为

$$l = f_{N_1}(x, \xi) = \sum_{i=1}^{N_1} \omega_{1i}^{(2)} f_s \left(\sum_{j=1}^N \omega_{ij}^{(1)} x_j + \theta_i^{(1)} \right) + \theta_1^{(2)} \quad (9)$$

其中 $f_s(I) = (1 + e^{-I})^{-1}$ 为 Sigmoid 函数; $f_{N_1}(x, \xi)$ 的下标 N_1 表示隐层神经元数, ξ 表示所有权和阈值参数构成的向量. Barron^[18] 指出, 网络所实现的函数与被逼近函数之间的最小均方逼近误差随 N_1 的增加按 $1/N_1$ 的速率下降

$$\int_{B_r} (\hat{f}(\mathbf{X}) - f_{N_1}(\mathbf{X}, \xi_0))^2 dP(\mathbf{X}) \leq \frac{(2rC_f)^2}{N_1} \quad (10)$$

其中 $C_{\hat{f}}$ 在 $\hat{f}(\mathbf{X})$ 为 S 阶可导 ($S \geq 1$) 时为有限值, 并假设 $\mathbf{X}$ 的定义域包含原点, 且处于一个半径为 r 的超球 B_r 之内. 因此 N_1 只需随 N 线性增加, 但比起小波网络, 网络学习过程需要更多的训练样本.

由于不需要太多的隐含层神经元即可获得好的人脸逼近效果, 但考虑到每一点对的匹配估计值 σ , 决定了对应重建三维点的位置可信度. 因此, 为获得更精确的网络逼近结果, 设置可信度越高的点参加越多次数数的训练过程. 具体的, 对于每一点对所反求的三维训练样本, 令其参加前 N 次训练循环过程

$$N = \lfloor C(\sigma - \epsilon)^2 \rfloor \quad (11)$$

其中, ϵ 为匹配拒判阈值, C 为与最大循环次数有关的常数. 在我们的系统中, 取 $\epsilon = 0.5$, $C = 100$.

对图 4(b) 进行重建, 效果见图 5. 当人脸曲面重建以后, 需要归一化到基准平面. 这个平面应由纹理丰富、容易在二维图像中精确定位并且对人脸表情、光照以及拍摄角度鲁棒性强的特征点确定. 在我们的系统中, 取双眼中心以及鼻孔连线的中心三点确定基准平面, 称这样的三点组成的三角形为归一化三角形. 建立三维坐标参照系, 如图 6 所示.

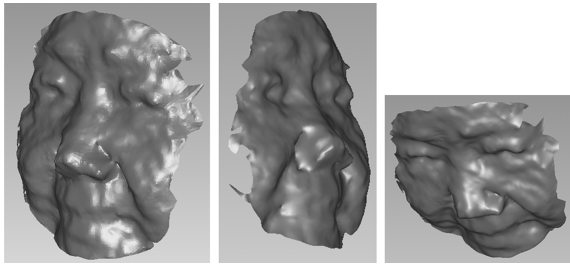


图 5 神经网络人脸曲面重建

Fig. 5 Face reconstruction by neural network

由于在实际采样中, 要求人的目光正视前方 (如图 6(a) 所示) 是苛刻的, 极端情况发生在图 6(b). 因此, 如图 6(c) 所示从两鼻孔中心向左右眼的连线作垂线, 交于 o 点, 建立直角坐标系, 使得在各种情况下都能获得较精确的归一化结果. 所以, 在图像采样时我们仍然要求头部不能过分倾斜, 至少需要如图 6(d) 中所示能够定位双眼位置. 此时, 取单侧鼻孔代替两鼻孔中心建立直角坐标系进行归一化处理. 但应该指出的是, 即使能够定位, 这样的归一化仍然是粗略的, 在后续的比对识别中还需进行迭代调整.

于是, 对于每张人脸曲面, 只需记录相应的神经网络参数以及归一化三角形作为三维人脸特征即可.

3.2 人脸识别

三维人脸的识别过程实际上即为归一化的两张神经网络曲面的相似度判别过程. 如果采集的是正面的人脸图像, 并且归一化足够精确, 则识别过程相

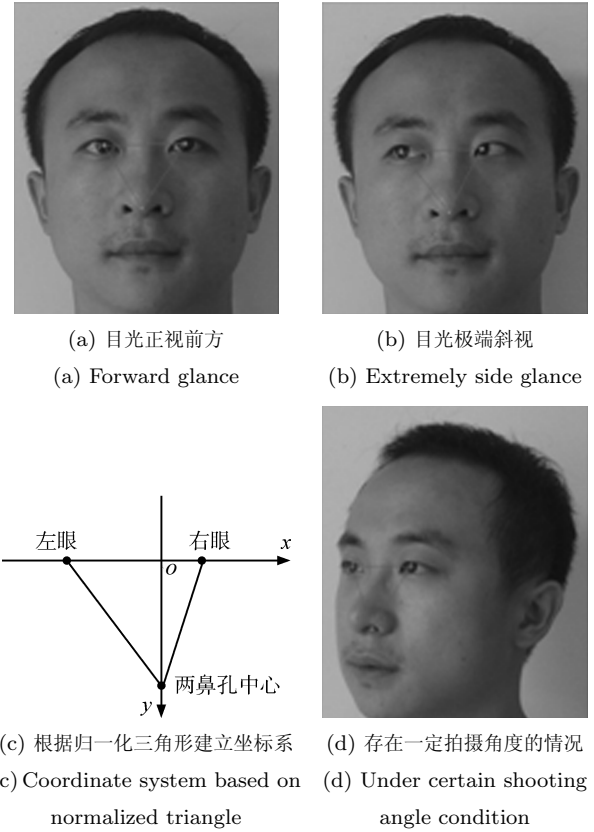


图 6 确定归一化三角形

Fig. 6 Obtaining normalized triangle

对容易. 整个过程分两步:

步骤 1. 比较瞳距以及鼻孔到左右眼连线的距离是否相近, 如差值在一定阈值范围内, 则说明两张曲面有可能是同一人, 需进行进一步的判别, 否则可以认为不属于同一人.

步骤 2. 假设需对比的两张曲面分别为 $z = f(x, y)$ 和 $z = g(x, y)$. 那么定义

$$d = \frac{1}{N} \sum_{(x,y) \in S} |f(x, y) - g(x, y)| \quad (12)$$

作为判别尺度, 其中 S 定义为与归一化三角形有关的椭圆, 而 N 为 S 的面积. 设在新的坐标系下, 图 5(c) 中的归一化三角形坐标分别为: $(a, 0, 0)$ 、 $(-b, 0, 0)$ 、 $(0, c, 0)$, 一般地, 定义 S 为如下椭圆区域

$$\frac{x^2}{\left(\frac{2.5(a+b)}{2}\right)^2} + \frac{\left(y - \frac{c}{2}\right)^2}{(2c)^2} \leq 1 \quad (13)$$

式 (12) 中, 如果 d 在一定的阈值内, 即可判断该两张曲面属于同一人.

然而, 在实际曲面归一化过程中, 由于鼻孔点的定位通常不十分准确, 并且, 在有些时候目光也存在

向上与向下的不确定性, 因此, 这样的定位总的来说是比较粗糙的. 故我们需要将两张曲面进行迭代对齐, 设有两张曲面 $z = f(x, y)$ 和 $z = g(x, y)$, 迭代过程如下:

步骤 1. 从曲面 $z = g(x, y)$ 中选取整数点, 计算其到曲面 $z = f(x, y)$ 的最小距离;

步骤 2. 运用 Levenberg-Marquardt 算法对旋转和平移参数进行非线性最小二乘拟合;

步骤 3. 将旋转平移参数运用到曲面 $z = g(x, y)$ 形成新的曲面, 仍记作 $z = g(x, y)$;

步骤 4. 重复步骤 1~3 直到收敛.

另外, 当采集非正面人脸图像时, 如图 7(a)、7(b) 所示, 归一化后只能得到如图 7(c) 所示的半边脸重建结果. 那么, 由于缺少左边脸部表面三维点云信息, 神经网络重建曲面在这一部分的形状随意性很强, 给识别过程带来困难. 因此, 将只进行右半区域以及眼睛部分的人脸比对. 如果比对的两张人脸曲面是从不同左右角度进行采集的, 那么由于人脸具有很好的对称性, 我们只需要将其中一张人脸曲面做关于 $y-z$ 轴的镜向, 如图 7(e) 所示, 再将两幅点云进行 ICP 迭代, 便可获得如图 7(f) 的结果. 从图 7(f) 中可以了解到, 由于拍摄角度原因, 被鼻子遮挡的脸颊部分以及脸的侧面被很好地恢复了. 以上方法也被运用于当人脸 (特别是颧骨部分) 皮肤过于细腻而缺少纹理时出现大孔时的填补过程.

当人脸库十分庞大时, 比对过程需考虑时间消费. 为提高比对效率, 可分层次对人脸进行神经网络重建. 即首先选用较少的隐含层神经元进行粗略比对, 排除掉库中大部分人脸, 再提高隐含层神经元个数来进行精确比对, 以获得更高的效率.

另外, 采样的是基本正面的人脸时, 可考虑对重建曲面进行左右对称性检测, 出现不对称的情况通常发生在人的表情十分丰富, 或者面颊部分非常光洁引起的匹配失败时. 在这些情况下, 应该只对眼部周围以及鼻子附近这些纹理信息丰富并且不容易随表情变化而变化的区域进行比对识别, 如图 8 所示的圆形区域.

4 系统实例与讨论

采用一台 P4 2.8 G, 512 Mb RAM 的 PC 机, 两台 SAMSUNG SCC-C4201P CCD 摄像机, 分辨率 480 线, 有效像素 $752(H) \times 582(V)$, 建立如图 1 所示的双目立体视觉系统作为前端人脸图像采集. 整个三维人脸识别系统主要由四部分组成: 前端人脸图像采集系统、人脸三维点云重建系统、人脸三维曲面重建系统和三维人脸识别系统. 其程序流程如图 9 (见下页) 所示.

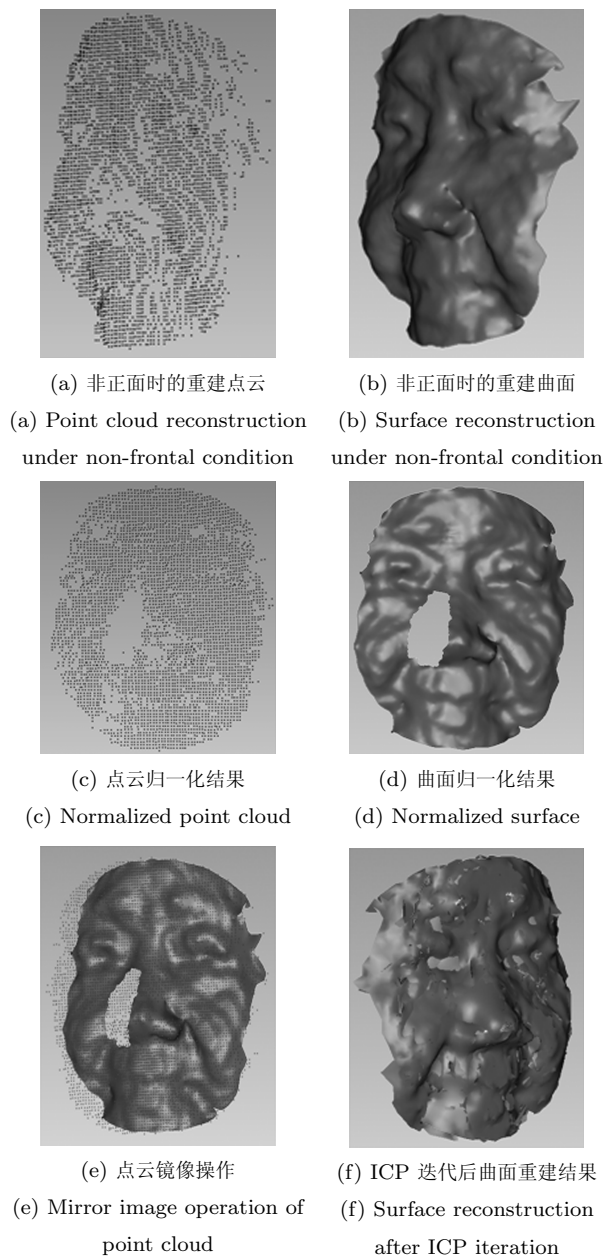


图 7 存在一定拍摄角度时的点云重建情况

Fig. 7 Reconstruction of 3D point cloud in different shooting angles

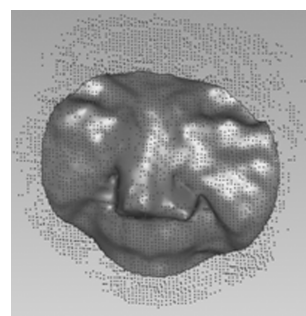


图 8 取人脸中心的圆型区域

Fig. 8 Circular region in the face center

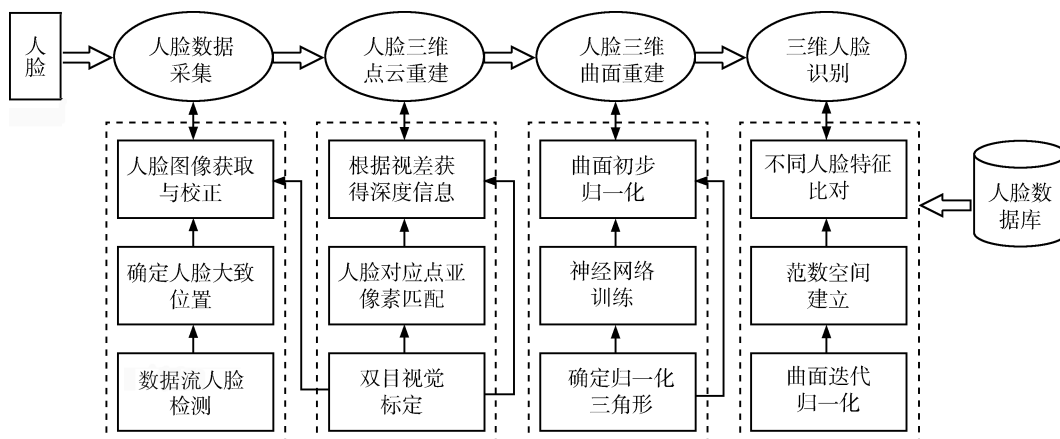


图 9 三维人脸识别系统流程

Fig. 9 Flow of the 3D face detection system

为了验证本文提出的三维人脸识别方法的有效性,我们在不同光照、背景条件下,在相隔半年的时间对多人进行多次不同姿态与表情的样本采集,为获得较理想的重建与识别效果,通常需要使人脸尽可能充满整个图像对。但在实际操作中,拍摄距离在 40cm~80cm 时,对重建精度的影响不大。建立如图 10 所示的三维人脸库。

本次实验主要选取了具有代表性的一些样本(如图 11 所示)进行较为细致的测试分析。其中样本 C1~C5 与样本 C6~D3 为亲姐妹,而样本 A1~A3 与样本 D8、样本 A4~B8 与样本 D4~D6 虽然是同一人,但前后采样时间间隔都在半年以上,并且各样本在采样时的光照、背景、姿态与表情都存在较大的差异。因此,对该组样本的分析具有一定的普遍性意义。

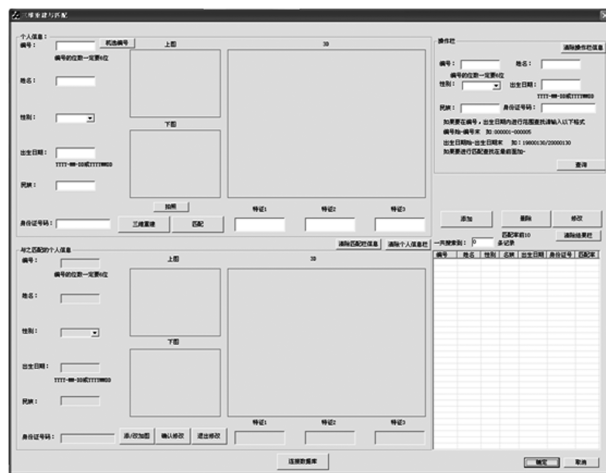


图 10 三维人脸库

Fig. 10 Database of 3D faces

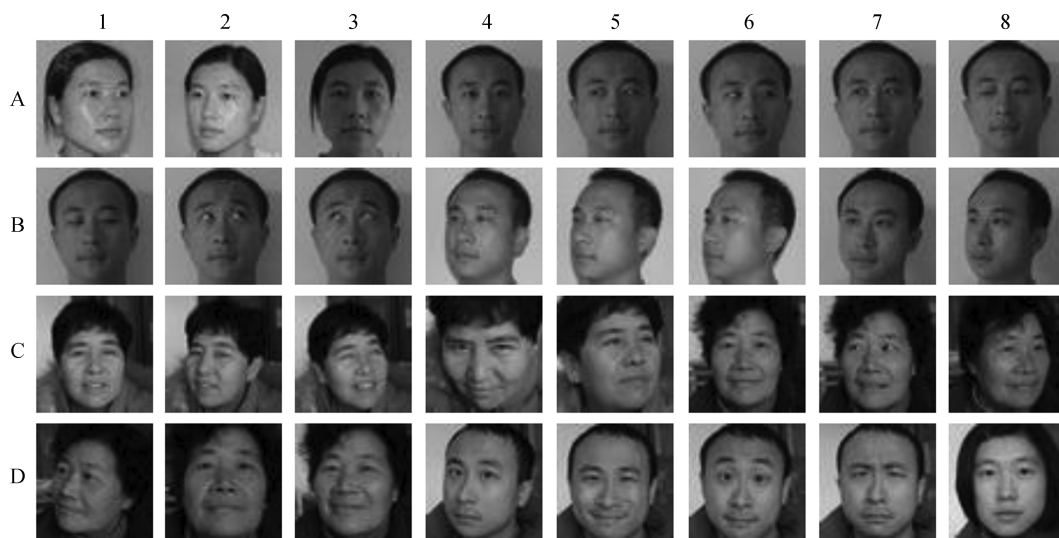


图 11 部分人脸样本

Fig. 11 Face samples

表 1 不同表情与拍摄角度下, 两人脸间的判别尺度

Table 1 Scale determination between faces in different expressions and shooting angles

-	C1	C2	C3	C4	C5	C6	C7	C8	D1	D2	D3
C1	—	0.854	1.342	1.682	1.663	17.494	13.393	15.106	13.338	15.409	14.257
C2	0.840	—	1.173	0.988	1.903	14.231	13.990	13.854	15.294	14.003	16.325
C3	1.324	1.576	—	2.001	1.435	15.346	14.245	13.865	16.132	14.447	16.235
C4	1.863	2.128	1.135	—	1.864	15.558	16.337	13.540	13.978	15.637	17.029
C5	1.647	0.437	1.753	2.012	—	16.452	15.196	16.346	13.890	15.569	14.591
C6	16.534	16.223	15.382	14.980	16.015	—	1.324	2.097	1.457	0.962	0.550
C7	15.346	16.479	14.890	17.344	15.345	1.239	—	2.091	1.189	2.003	1.769
C8	15.782	15.890	17.129	17.912	15.373	1.139	2.261	—	2.266	1.198	1.984
D1	16.498	14.598	16.676	17.120	15.337	1.357	2.159	1.112	—	2.490	3.109
D2	15.256	15.257	17.863	16.660	15.589	0.998	2.228	1.678	1.870	—	1.669
D3	16.367	14.509	16.712	15.893	16.739	1.809	1.294	2.227	1.749	2.119	—

对图 11 中样本进行自动归一化三角形定位, 准确率达到 94%, 其中只有样本 B1 与 B6 出现瞳孔难以定位的情况. 对其进行交互瞳孔定位后, B1 能完成归一化, 但 B6 由于脸部拍摄角度过大, 造成归一化过程失败.

对成功归一化的样本重构后, 进行两两相似度比对, 计算式 (12) 中所确定的判别尺度 d , 可以发现, 对于不同人之间无论拍摄情况如何多变, 其判别尺度 d 相对都很大, 即使对于亲姐妹的 C1~C5 与 C6~D3 (如表 1 所示), 也能很容易地进行区分. 并且从表 1 中不难看出, 表情的变化对识别过程影响不大, 这是由于我们采用了对表情变化相对不敏感的圆形区域 (如图 8 所示) 进行比对识别. 但在表情变化十分剧烈时, 仍然会遇到一些困惑. 如样本 D5 与 D6 的判别尺度为 5.967, 虽然仍具有一定的聚类性, 但已使得判断变得模糊. 而对于不同的光照情况, 如样本 A1~A3, 对识别效果没有影响. 在分类阈值设定为 7 的条件下, 样本都能正确聚类. 在本次实验中, 总体识别率达到 99%.

实验表明, 系统允许接受非正面的人脸采集图像, 只要在该图像中能方便地识别出归一化三角形即可. 系统识别基本不受光照条件的影响, 对不同面部表情有一定的鲁棒性, 并且对不同时期不同发型的人脸也有较好的识别准确度. 在匹配算法中, 每一点都可并行计算, 因此在实际编程中, 采用多线程的处理方式可在很大程度上提高算法的速度. 在本次实验中, 采用双线程, 对每张脸计算 10 000 个匹配点, 用时大约在 20 秒左右. 而随着计算机硬件技术的提高, 以及采用多 CPU 的方式, 有望使立体视觉重建过程实时化.

5 结论

本文提出了一种新的基于双目立体视觉采集人脸图像, 运用复小波变换技术精确匹配对应点, 通过神经网络方法实现人脸重构的三维人脸识别方法, 与传统使用激光扫描仪或结构光等人脸信息采集方法相比, 由于使用了非侵入式的双目立体视觉系统, 使采集过程更加友好、快捷, 并可根据需要达到隐蔽采集的目的. 本文得出了以下结论:

1) 运用基于复小波变换的相位相关算法可获得密集并且满足识别精度要求的人脸三维点云信息. 该算法对背景无特别要求.

2) 通过三维点位置可信度参数 σ 决定其在神经网络训练次数的方法, 使得重建曲面更加精准. 而归一化处理使得识别过程为后继的迭代识别比对提供较理想的初始状态.

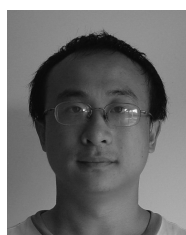
3) 采用局部人脸比对的思想, 在表情剧烈或者面颊特别光洁的情况下, 只对纹理丰富并且相对不受表情变化影响的双眼以及鼻子周围区域进行重建比对, 减少了拒判的概率.

研究表明, 该方法不仅对环境以及人脸位姿表情等鲁棒性强, 识别准确率高, 并且成本低廉, 适合在许多领域推广应用.

References

- 1 Zhao W, Chellappa R, Phillips P J, Rosenfeld A. Face recognition: a literature survey. *ACM Computing Surveys*, 2003, **35**(4): 399–458
- 2 Phillips P J, Wechsler H, Huang J, Rauss P. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing Journal*, 1998, **16**(5): 295–306
- 3 Bronstein A M, Bronstein M M, Kimmel R. Three-dimensional face recognition. *International Journal of Com-*

- puter Vision, 2005, **64**(1): 5–30
- 4 Gokberk B, Salah A A, Akarun L. Rank-based decision fusion for 3D shape-based face recognition. In: Proceedings of the 13th Signal Processing and Communication Applications Conference. New York, USA: IEEE, 2005. 1019–1028
 - 5 Lee Y, Song H, Yang U, Shin H, Sohn K. Local feature based 3D face recognition. In: Proceedings of the 5th International Conference on Audio- and Video-based Biometric Person Authentication. New York, USA: Springer, 2005. 909–918
 - 6 Lu X G, Jain A K. Deformation analysis for 3D face matching. In: Proceedings of the 7th IEEE Workshops on Applications of Computer Vision. Breckenridge, Colorado, USA: IEEE, 2005. 99–104
 - 7 Russ T D, Koch M W, Little C Q. A 2D range Hausdorff approach for 3D face recognition. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 169
 - 8 Pan G, Han S, Wu Z H, Wang Y M. 3D face recognition using mapped depth images. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 175
 - 9 Chang K I, Bowyer K W, Flynn P J. Adaptive rigid multi-region selection for handling expression variation in 3D face recognition. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 157
 - 10 Passalis G, Kakadiaris I A, Theoharis T, Toderici G, Murta N. Evaluation of 3D face recognition in the presence of facial expressions: an annotated deformable model approach. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 171
 - 11 Ansart A N, Abdel M M. 3D face modeling using two orthogonal views and a generic face model. In: Proceedings of International Conference on Multimedia and Expo. Baltimore, Maryland, USA: IEEE, 2003. 289–292
 - 12 Tang L A, Huang T S. Automatic construction of 3D human face models based on 2D images. In: Proceedings of International Conference on Image Processing. Lausanne, Switzerland: IEEE, 1996. 467–470
 - 13 Zhang Y, Prakash E C, Sung E. Hierarchical modeling of a personalized face for realistic expression animation. In: Proceedings of the IEEE International Conference on Multimedia and Expo. Lausanne, Switzerland: IEEE, 2002. 457–460
 - 14 Zhang Z Y. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(11): 1330–1334
 - 15 Zhang Z Y. Flexible camera calibration by viewing a plane from unknown orientations. In: Proceedings of the IEEE International Conference on Computer Vision. Kerkyra, Greece: IEEE, 1999. 666–673
 - 16 Lienhart R, Maydt J. An extended set of Haarlike features for rapid object detection. In: Proceedings of the IEEE International Conference on Image Processing. New York, USA: IEEE, 2002. 900–903
 - 17 Lienhart R, Kuranov A, Pisarevsky V. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In: Proceedings of the 25th Pattern Recognition Symposium. Berlin, German: Springer-Verlag, 2003. 297–304
 - 18 Barron A R. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*, 1993, **39**(3): 930–945



周佳立 浙江大学 CAD & CG 国家重点实验室博士研究生. 主要研究方向为图形图像处理、计算机视觉、模式识别. 本文通信作者.

E-mail: zhoulie@gmail.com

(**ZHOU Jia-Li** Ph.D. candidate at the State Key Laboratory of CAD & CG, Zhejiang University. His

research interest covers graphics and image processing, computer vision, and pattern recognition. Corresponding author of this paper.)



张树有 教授, 博士. 主要研究方向为产品数字化设计和计算机图形学.

E-mail: zsy@zju.edu.cn

(**ZHANG Shu-You** Professor, Ph.D.. His research interest covers product digitalization design and computer graphics.)



杨国平 浙江中医药大学信息技术学院副教授, 主要研究方向为图像处理.

E-mail: ygp@zjtcn.net

(**YANG Guo-Ping** Associate professor at the College of Information Technology, Zhejiang Chinese Medical University. His main research interest is image processing.)