

基于双目视觉的立体人脸识别系统^①

Stereo Face Recognition System Based on Binocular Vision

吴翔 唐焕华 刘锦高 (华东师范大学 电子科学与技术系 上海 200241)

摘要: 提出了一种基于立体视觉的人脸识别算法及介绍其系统实现。在传统二维人脸识别的基础上, 根据人眼视觉原理设计双摄像头提取人脸的三维信息作为判别标准, 增加了识别的准确度和鲁棒性。对于左右两个摄像头的输入视频帧, 依据相关系数和约束关系找到两幅图每点的对应关系, 进而得到同一个 3D 空间点投影到不同摄像机图像上对应点的位置差, 由此表征出深度信息。这些 2D 图像不具备的深度数据组成不同传统人脸识别的新判别依据。还将给出实现本算法的系统构成及最后的实验结果, 在复杂环境下试验证明三维的人脸识别具有更好的鲁棒性和更高的识别率。

关键词: 人脸识别 景深 三维空间 相关系数 视差

1 引言

本文利用立体视觉传感器确定 3D 点在空间中的位置^[1], 通过几何与代数知识计算视差, 恢复场景的深度信息。再把生成的三维深度图像, 应用于人脸识别研究。经过三维数据校准、匹配, 最终获得一个基本不受光照与姿态影响, 识别率明显优于传统基于二维图像人脸识别的新型系统。

研究包括立体视觉传感器的设计, 确定图像匹配点, 选择合适的约束条件, 精确计算 3D 空间景深信息^[2], 构建三维的人脸模型, 及最后的三维匹配与数据库快速搜索算法^[3]。本文组织如下: 第二部分介绍系统整体构架, 硬件模块框图和系统实现在这章给出。第三部分提出有效区域分割和匹配对应点的算法。三维信息的计算以及如何用来进行人脸识别的方法在第四部分进行探讨。第五部分是实验的结果, 最后, 第六部分总结全文并对未来工作展望。

2 系统硬件结构

系统硬件结构设计与系统实际应用需求密切相关, 本系统的整体框图如图 1 所示, 可分为三个部分。这里给出宏观的功能描述。

作为身份识别依据的立体人脸图像对, 通过水平

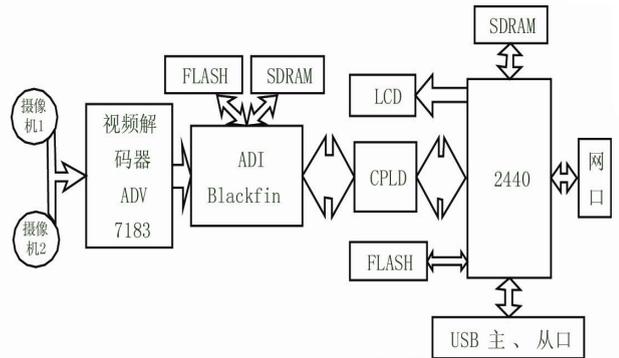


图 1 立体人脸识别系统框图

放置并间隔 20cm 的双摄像头获取^[4]。模拟信号经视频解码器进行数字化和编码, 得到的数据输入 ADI 的 DSP 处理^[5]。ARM 2440 及其外设用来产生用户的操作界面, 并且提供用户输入命令的平台。并且使用 ARM 模块可以方便实现 LCD 驱动及 USB、以太网等高速通信。两者间的 CPLD 模块主要完成硬件控制、协调及缓冲的功能, 如果需要还可以分担 DSP 的部分运算压力。

3 匹配对应点算法

可由简单的几何知识确定 3D 点在空间中的位置。使两台摄像机 L 和 R 的 X 轴重合, Y 轴和 Z 轴分别相

^① 基金项目:上海市 2007 年科技攻关重点项目(075115002);华东师范大学优秀博士研究培养基金项目(20080050)

收稿时间:2008-10-06

互平行。Y 轴垂直于纸面所以在实际推导中并不使用。右侧摄像机的原点或投影中心的偏移量为 b 。目标点 P 在左图像中对应点为 P_l ，在右图像中对应点为 P_r 。具体如图 2 所示。显然 P 点位于光线 $L P_l$ 和 $R P_r$ 的交点处。根据相似三角形可推出 $z = fb/(x_l - x_r) = fb/d$ 。

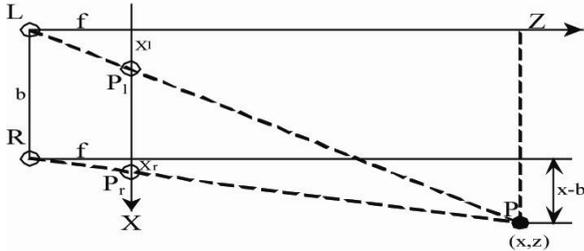


图 2 立体视觉系统几何模型. b 是基线

可见 P 点的深度和摄像机的焦距 f ，基线长度 b 以及视差 $d = x_l - x_r$ 有关^[6]。而 f 和 b 在硬件确定后是固定的，也就是通过两幅图上表征同一点的 x 方向坐标差 d ，可得到实际对应点的深度。接下来我们详细讨论对于左图中的一点，如何知道右图中哪一点和它对应。

3.1 有效区域分割

由于两个摄像机位置不一样，所以两幅图像背景也不一样，要分离出有效的区域进行比较必不可少。一对由双目摄像头拍摄的照片如图 3(a)、3(b)所示，其中图 3(a)是右摄像头所拍而图 3(b)是左摄像头拍摄。

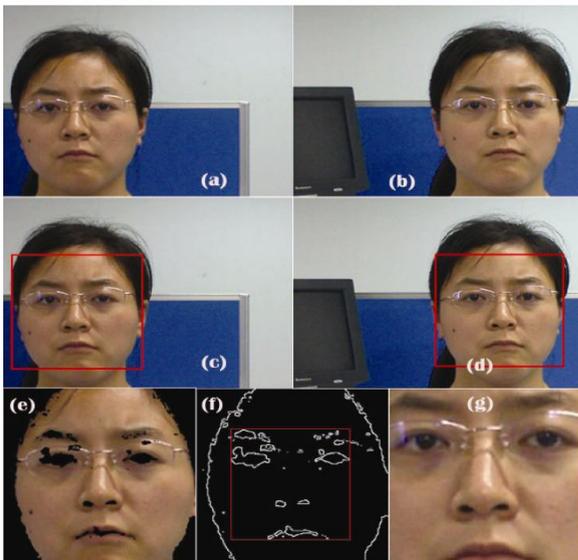


图 3 (a)右眼图 (b)左眼图 (c)右图人脸追踪结果 (d)左图人脸追踪结果 (e)肤色检测结果 (f)边缘检测结果 (g)三维分析区域

可采用二维人脸识别中比较成熟的 Adaboost 分类器方法进行人脸背景分离。本系统使用 Intel 的开源计算机视觉库 OpenCV，将其移植到 ADI 的 DSP 中得以实现。调用其中的 `cvLoadHaarClassifierCascade` 函数可以获得一个级联的 boosted 分类器，再调用 `cvHaarDetectObjects` 函数则可利用这个分类器在背景中找到人脸。人脸追踪结果如图 3 (c)和(d)。

对于人脸区域，进一步要划分出具备显著深度信息的五官区域。先利用肤色检测，采用人脸皮肤颜色模型 $C_b \in [86, 140], C_r \in [130, 168], Hue \in [0.01, 0.1]$ 得到如图 3(e)所示的结果，在此基础上用 `canny` 算子进行边缘检测，可以获得图 3(f)中的五官边缘，最终分割出需要三维分析的区域图 3(g)。

3.2 搜索对应点

随机考察图 3(g)(右眼图)中一个点，位置在第 100 行 60 列。第 100 行的像素亮度-坐标图如图 4(a)所示。一般来说亮度最高的点如 93 点左右为鼻子上的点。这里要求出另外一幅图即左眼图上和该点对应点的坐标。

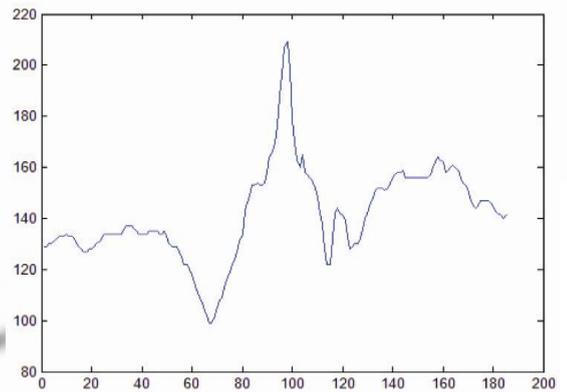


图 4 (a) 单行像素亮度坐标图

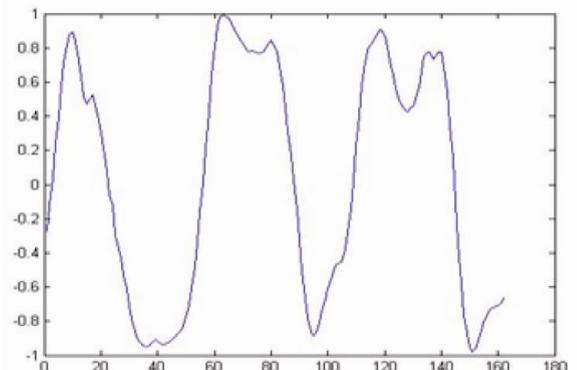


图 4 (b) 右图某点和左图各点的相关系数

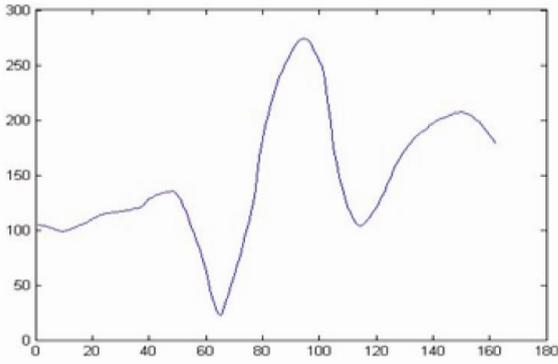


图 4 (c)和左图各点的欧式距离

根据外极线约束的准则，由于摄像头处于相同高度且水平，我们认为对应点也应该出现在左眼图的第 100 行上。扫描左眼图第 100 行的每个像素亮度值，分别计算它们和右眼图中第 100 行 60 列的点及其附近点的相关系数^[7]：

$$Corr(l,r) = \frac{Cov(\vec{f}_l, \vec{f}_r)}{sd(\vec{f}_l)sd(\vec{f}_r)}$$

得到图 4(b)。从 4(b)中可以看出，左眼图中第 14 列、63 列、118 列及 137、143 列的点和右眼图中第 60 列的点区域的相关系数都比较大。同样的方法再考虑欧式距离，可绘出图 4(c)，可见只有 63 列的点在相关系数很大的同时又和原点的像素亮度具有很小的欧式距离，所以为所求的对应点。该点视差即为 63-60=3pixel。用这种方法可以求出待分析区域内所有点的视差^[8]。

4 三维匹配

如上节所述，可以获得具备人脸显著凹凸信息的五官区域各点的视差值。进行人脸三维重构，就是用这些视差表征出面部各点的深度。

不失一般性，这里考察最亮点，即第 54 行和 100 行中处于鼻子位置的点。先看第 54 行的情况。在图 5(c)中，最亮点出现在 x=96 处，而在图 5(d)中，最亮点在 x=98 的位置，视差为 2。在第 100 行的点中，图 5(e)最大值处于 x=98，图 5f 中最大值处于 x=103，视差为 5。可见，鼻子在第 100 行上的点比在第 54 行中的点投影在双目摄像头上的视差要大。

根据对图 2 所示几何模型的分析，意味着前者离摄像头平面更近。事实上，由图 5(a)和(b)可见第 100

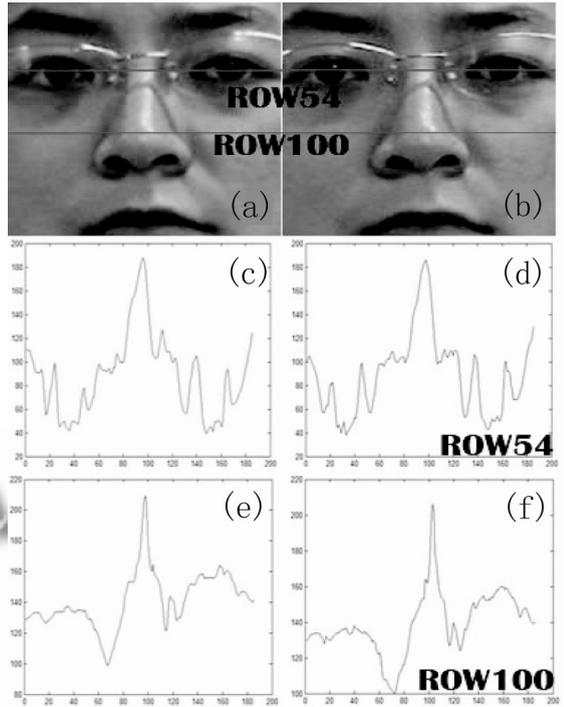


图 5 (a)、(b)分别为右眼和左眼观测五官区域，(c)、(d)分别是右眼和左眼图第 54 行像素亮度-坐标图，(e)、(f) 分别为右眼图和左眼图第 100 行像素亮度-坐标图

行正处于鼻尖的位置，鼻尖点比第 54 行上鼻子的点更接近摄像头是和实际情况完全吻合的。

同理，用以上方法推得所有关键区域点的景深，用这个数据替代各点的亮度进行匹配，这是本文提出方法和传统二维人脸识别最大的不同。前者包含了三维的信息，可以减少对光线和姿态的依赖性。当然，获得三维景深信息的同时也可以轻易获得各点亮度信息，所以二维人脸识别所运用的各种方法同样在这里也是适用的。实际上传统识别算法中的人脸定位、提取，及识别中的 PCA^[9]等算法都可以用于立体人脸识别，其区别就是用深度信息替代亮度信息，实现基于 3D 算法结合 2D 算法，互补长短既可以解决实用性差和算法复杂度高的问题，又可以获得优良的识别效果。

5 实验结果

图 6 是所使用的实验平台，具体结构已在图 1 中介绍。获得的数据进行三维重构达到描绘人类深度的目的，如图 7 所示。

同时另外一个验证实验使用一张彩色打印人脸的纸，试图欺骗识别系统。由于肤色和二维亮度信息的



图6 实验平台



图7 三维匹配结果

分布完全符合真实人脸情况，传统人脸识别算法无法分辨这种伪装。但是在本文提出的系统下，所有五官点的视差测得都基本相同，即面部点处于同一平面，所以可以被系统立刻识别。另外对于正常识别情况，在二维识别的基础上融入三维信息也大大提高了识别率和对光线和背景噪声的鲁棒性。

6 结论

本文根据人眼双目成像的原理，通过双摄像头实现获得立体信息进而提出人脸识别的一种新方法，并给出实现系统的结构。相比传统的人脸识别方法，由于本算法考虑了立体信息，不仅仅依赖于受光线影响严重的亮度数据，所以无论在鲁棒性和识别率上都有很大的提高。而专用的全息扫描获得三维数据的方法^[10]，设备昂贵且采样非常不方便，不如本文提出方法耗材简单且取样非常方便。

同时值得指出的是，三维的人脸模型的匹配和对立体图像对的数字处理都是非常耗费时间的^[11]，本文研究的算法在嵌入式系统中实现。虽然使用的数字处理芯片 ADI Blackfin 主频可高达 600MHz，但是为了保证模式识别的实时性，高效、快捷、节约的优化算法需要进一步研究。另外由于获得数字信号即在对连续信号采样时有损失，所以要获得高精度的视差数据需要高分辨率的摄像头。更高分辨率的摄像头意味着

更好的效果也同时意味着更复杂的运算量，如何既可以克服实用性差和算法复杂度高的困难，又可以获得优良的识别效果，这也是未来需要进一步研究的问题。

参考文献

- 1 Griesser A, Koninckx TP, Gool LV. Adaptive real-time 3D acquisition and contour tracking within a multiple structured light system. Proceeding of the 12th Pacific Conference on Computer Graphics and Applications, October 2004:361 - 370.
- 2 Smolic A, Kauff P. Interactive 3-D video representation and coding technologies. IEEE, 2005,93(1):98 - 110.
- 3 Bleyer M, Gelautz M. A layered stereo matching algorithm using image segmentation and global visibility constraints, Photogrammetry Remote Sensing, 2005 (59):128 - 150.
- 4 Torr PHS, Murray DW. Statistical detection of independent movement of a moving camera. Image Vision Compute, 1993,1(14).
- 5 陈峰. Blackfin 系列 DSP 原理与系统设计. 北京: 电子工业出版社, 2004:57 - 151.
- 6 Shapiro LG, Stockman GC. 计算机视觉. 北京: 机械工业出版社, 2005:293 - 294.
- 7 Kolmogorov V, Zabih R. Computing visual correspondence with occlusions using graph cuts. Proceedings of International Conference on Computer Vision, 2001.
- 8 Sanger TD. Stereo disparity computation using Gabor filters. Biol. Cyber, 1988(59):405 - 418.
- 9 Turk M, Penland A. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 1991:371 - 86.
- 10 Blanz V, Vetter T. Face recognition based on fitting a 3D morphable model. IEEE Trans. Pattern Anal. Mach. Intell, 2003,25(9):1063 - 1074.
- 11 Chiuso A, Favaro P, Soatto JH.S: Structure from motion causally integrated over time. IEEE Trans. Pattern Anal. Mach. Intell, 2002,24:523 - 535.