

虚拟环境中的人脸姿态交互方法

杭 飞 刘惠义

(河海大学 计算机与信息学院, 南京 210098)

摘 要 提出一种基于面积比的人脸姿态估计方法,先分析人脸姿态发生变化时特征点之间形成的三角形的面积变化,再应用BP神经网络对位置参数和人脸姿态参数的关系进行学习,从而对人脸姿态进行估计,最后将该方法应用于虚拟环境的漫游中。实验结果表明,采用该方法对人脸转动进行估计,采用的特征点比较少,具有较高的识别率和稳定性。

关键词 姿态估计 神经网络 人机交互

中图法分类号: TP391.4 文献标志码: A 文章编号: 1006-8961(2010)08-1145-06

A New Method for Face Pose Interaction in Virtual Environment

HANG Fei, LIU Huiyi

(Computer & Information Engineering College, HoHai University, Nanjing 210098)

Abstract A new method is proposed for face pose estimation based on the area ratio in the paper. First, the change of triangle area ratio which consist of feature points is analyzed when a face turns. Then, the BP neural network is applied to train the relationship between position parameters of area ratio of face and the parameters of pose for face pose estimation. Finally, the method is applied to roam in the virtual environment. The experimental results demonstrate that this method can estimate accurately pose using only few characteristic points, and also does well in recognition rate and stability.

Keywords pose estimation, artificial neural network, human-computer interface

0 引 言

传统的人机交互是以键盘、鼠标等设备与计算机进行交互。近年来,随着计算机计算能力的改善和交互设备性能的提高,以用户为导向,自然、高效的多通道人机交互技术成为人机交互领域研究与应用的热点^[1]。

人脸姿态是视频图像中一个很重要的交互特征,是姿态估计问题的一个具体化分支。人脸姿态估计就是在2维图片或者视频序列流的某一时刻内,近似地计算出图片或视频中人脸对应与 X, Y, Z 3个空间坐标轴的偏转角度^[2]。具体地说,人脸姿态的变化包括绕 X, Y, Z 轴的旋转和沿 X, Y, Z 轴的

平移这6个自由度(其中,绕 X, Y 和 Z 轴的旋转分别称为Tilt, Yaw和Roll方向旋转,如图1所示),人脸姿态估计就是指确定输入图像中这6个自由度的过程。人脸姿态估计作为计算机视觉中的一项基本课题,在人脸识别、计算机视觉、自动视频监控、表情判断、智能人机交互、司机疲劳检测等方面都有广泛

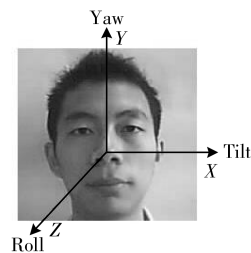


图1 人脸的3个旋转方向

Fig. 1 3 Directions of rotation of face

基金项目:国家自然科学基金项目(60673186)

收稿日期:2010-03-29;改回日期:2010-05-10

第一作者简介:杭飞(1984—),男,2009年于河海大学获模式识别与智能系统专业工学硕士学位。主要研究方向为人机交互。

E-mail: hangflying@gmail.com

的应用前景。

目前,国内外的研究人员提出了很多用于人脸姿态的估计方法,这些方法大体上可以分为基于人脸特征的方法和基于模型的方法^[3-4]。现有的人脸姿态估计方法中,大多属于基于模型的方法。Nikolaidis, Pitas^[5]采用由人脸图像中的 3 个特征点(两个眼睛和嘴巴)确定人脸姿态,该方法对眼睛和嘴位置的精确度非常敏感。Ballard^[6]采用 3 个特征点进行姿态估计,前两个点为眼睛中心,后一个点为鼻尖或嘴角,通过循环算法来求解。Ho 等人^[7]提出选取两个嘴角、两个眼角,采用透视投影进行姿态估计。Horprasert 等人^[8]采用两个外眼角、两个内眼角、鼻尖这 5 个特征点建模人脸,他们利用人脸结构的统计模型和交比投影的不变性来进行姿态估计。该方法需要统计学中所获得的鼻梁和眼睛开口的长度。Smith 等人^[9]首先构建了头部的边框,然后利用眼睛和嘴角与脸部边界的距离来进行姿态估计。Hogg 等人^[10]采用协作人工神经网络学习算法将姿态旋转特征进行映射求解。Wei 等人^[11]采用基于特征空间的方法,引入 Gabor 滤波器来消除不同人脸形状的差异以获得更精确的估计。

上述基于模型的方法大多受到焦距准确性、人脸模型近似性的影响,有的还需要正视图中鼻子长度或嘴巴长度才能进行 Tilt 方向旋转的姿态估计。对此,提出一种基于面积比的人脸姿态估计方法。该方法利用了类似于三角模型算法,是一种基于几何模型的姿态估计方法,该方法利用两个眼睛、嘴巴、鼻子 4 个特征点在人脸中的连线形成的三角形面积比来估算出人脸绕 X 轴和 Y 轴的复合转动角度。该方法可以单纯靠一张图片定位出人脸的角度,与焦距无关,也不需要正视图中的鼻子、嘴巴的长度。而对不同的人脸模型存在的差异可以通过大量样本的神经网络训练逼近一个最优值。

1 人脸检测与特征点定位

1.1 人脸图像预处理

为了更精确地定位特征点,通常需要对图像进行预处理,包括图像平滑滤波、图像灰度变换和光线补偿。通过这样的处理后,图片中的人脸信息得到了加强,为后期特征点的识别工作打下了基础,更容易实现计算机自动检测和标定。

1) 图像平滑滤波

首先采用 Gaussian 函数平滑方法尽量消减图像中噪声的影响,其函数形式如下式

$$G_{\delta}(x, y) = \frac{1}{2\pi\delta^2} e^{-\frac{x^2+y^2}{2\delta^2}} \quad (1)$$

在充分考虑平滑范围以及对图像边缘信息的影响之后,合理地设置好参数 δ , 使用 Gaussian 函数对输入图像 $I(x, y)$ 进行平滑,输出图像 $\hat{I}(x, y)$ 其表示形式如下式

$$\hat{I}(x, y) = I(x, y) * G_{\delta}(x, y) = \iint I(x, y) \times G_{\delta}(x - m, y - n) dm dn \quad (2)$$

2) 图像灰度变换

图像的灰度变换是指根据目标条件按一定变换关系逐点改变原图像中每一个像素灰度值的方法。目的是为了改善画质,使图像的显示效果更加清晰,图像特征更为明显。根据面部图像的直方图曲线,可以找出规律并通过灰度变换突出图像的边缘特征,更容易地定位特征点。

3) 光线补偿

原始图像中的肤色信息比较容易受到光源、摄像头的色偏等影响,从而在整体上发生色彩的偏移(即通常所说的色彩偏冷、偏暖、偏黄等),为了抵消该影响,按亮度从高到低的顺序,将整个图像中所有像素进行排列,取前 5% 的像素^[12],如果占亮度前 5% 的这些像素的数目超过 100,就将它们的 R, G, B 分量色彩值都调整为 MAX(这里是 255)。

1.2 人脸检测

采用 OpenCV^[13]中基于灰度图像中弱特征检测方法对人脸进行检测定位。使用 OpenCV 中提供的一个 20×20 的“haarcascade_frontalface_alt.xml”分类器,用 cvLoad 函数载入该 XML 分类器后,对图像进行强制类型转换。OpenCV 中提供的用于检测图像中目标的函数是 cvHaarDetectObjects(), 该函数使用针对目标物体训练的级联分类器在图像中找到包含目标物体的矩形区域,并将这些区域作为一个包含若干矩形的 CvSeq 型可动态增长元素序列返回。函数通过不同的比例大小扫描窗口对图像进行搜索,返回的元素序列中每一个矩形代表一个可能的人脸区域。

1.3 面部特征点定位

在已经检测到的人脸区域范围内,进行人脸特征点的定位操作。

为了准确定位面部特征,需要对输入图像进行肤色分割,即要把输入实验图像变为二值化图像,该图像仅含有肤色像素点和非肤色像素点。图 2(a)为输入图像的人脸定位,(b)为相似度分布图,(c)为人脸区域分割后的二值化图,(d)为人脸边缘检测图。

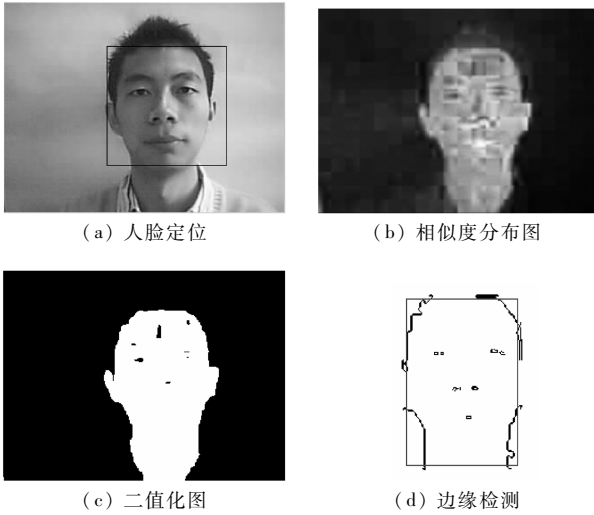


图 2 人脸定位与肤色分割

Fig.2 Face locating and skin color segmentation

对二值化图采用高斯拉普拉斯算子进行卷积计算,得到人脸区域的边沿检测图,在边沿检测图上半部分一定的区域对水平 X 轴方向进行投影,以便确定双眼在水平轴上的范围 A 和 B ,如图 3(a)所示。在 A 和 B 的上方区域对垂直 Y 轴方向进行投影,得到第一个峰值,如图 3(b)。设其附近的区域为 I , II 。 A 与 I 确定一个矩形区域 R_1 , B 与 II 也确定一个矩形区域 R_2 , R_1, R_2 即为眼睛所在的区域, R_1, R_2 的中心分别为双眼的中心。

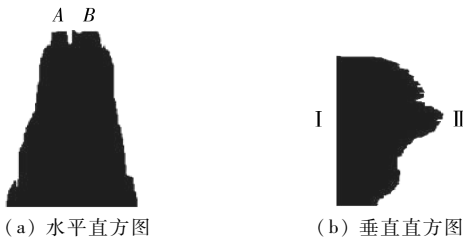


图 3 直方图

Fig.3 Histogram

由面部几何关系和五官的规则性,假设鼻子处在两个瞳孔之间的范围内。正视状态时鼻子一般位于人脸的中央,据此可以大致确定鼻子所处的水平区域和垂直区域范围。先确定鼻子区域像素值,然

后确定鼻子矩形区域,再次寻找鼻子水平区域最外侧的两点,两个坐标相加除以 2 就可以得到鼻尖点横坐标,用同样的方法可以得到鼻尖纵坐标。

嘴所特有的唇色和肤色有一定的差别,由于嘴唇的颜色更加集中,可以像建立肤色模型一样建立唇色模型^[14]。在眼睛定位后,可在眼睛下方的人脸区域内利用唇色模型对此区域进行唇色分割,得到一个唇色的连通域。再用类似眼睛定位的方法就得到嘴巴所在的区域。

2 基于面积比的人脸姿态估计方法

2.1 特征三角形的选取

我们发现,随着人脸姿态的变化,眼睛、鼻子和嘴巴在人脸中的相对位置也发生规律性的变化,它们形成的三角形面积也发生相应的变化。

如图 4(a)所示,点 A, B, C, D 分别代表左眼、右眼、鼻子、嘴巴所在矩形区域的中心点,连接 AB, AC, BC, AD, BD, CD 。延长 DC ,使 DC 的延长线交 AB 于 E 点。设 $m = S_{\triangle ADE} : S_{\triangle BDE}; n = S_{\triangle ABC} : (S_{\triangle ACD} + S_{\triangle BCD})$, S 为面积。

根据人脸的几何特征,可以得到在人脸正视前方的时候, $m = 1$ 。

如图 4(b)所示,在 Yaw 方向,当人脸向右转动时, m 逐渐减小;反之,当人脸向左转动时, m 逐渐增大。

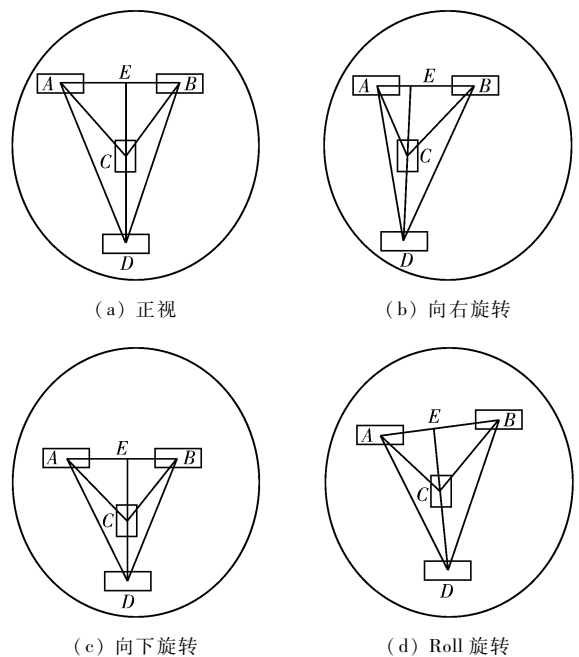


图 4 特征三角形模型

Fig.4 Characteristic triangle models

如图 4(c) 所示,在 Tilt 方向,当人脸向下转动时, n 逐渐增大;反之,当人脸向上转动时, n 逐渐减小。

如图 4(d) 所示,在 Roll 方向,当人脸绕 Z 轴旋转时, AB 两点的连线和水平线所成的角度就是人脸沿 Roll 方向旋转的角度。

设 θ_{Yaw} 为人脸在 Yaw 方向上的旋转角度, θ_{Tilt} 为人脸在 Tilt 方向上的旋转角度,可以运用 BP 神经网络建立有限个位置参数 m, n 和 $\theta_{\text{Yaw}}, \theta_{\text{Tilt}}$ 之间的对应关系。

2.2 运用 BP 神经网络进行训练

选择包含一个隐层的 3 层人工神经网络建立位置参数和姿态参数间的对应关系,先对人脸仅在 Yaw 方向上转动或者仅在 Tilt 方向上转动进行网络设计与姿态估计。

先讨论仅对 Yaw 方向上转动时的情况。

单一转动网络的输入层和输出层都只包含一个神经元,分别对应位置参数 m 和人脸绕 Yaw 方向转动角度 θ_{Yaw} 。通过增加隐层神经元的数目来提高该网络的精度,要比通过增加隐层数目的方法更加容易实现,适当增加几个隐层神经元来加速误差函数的收敛,在隐层神经元个数的选择上,采用 5 个神经元。初始权值选择 $[-1, 1]$ 之间的随机数。网络学习步长设定为 0.5,惯性系数设定为 0.3^[15]。

仅在 Tilt 方向上单一转动姿态估计的网络结构类似于 Yaw 方向。

当进行 Yaw 方向和 Tilt 方向的复合转动姿态估计的神经网络设计时,输入层有 2 个神经元,分别表示位置参数 m 和 n ,在隐层神经元个数的选择上,利用经验我们用 20 个神经元;输出层采用 2 个神经元,分别表示 θ_{Yaw} 和 θ_{Tilt} 两个姿态参数。

2.3 实验结果

从实际情况考虑,限定三个角度的偏转都在 $[-50^\circ, 50^\circ]$ 之间,规定:在 Yaw 方向上,人脸向左转动时 θ_{Yaw} 为负值,向右转动时 θ_{Yaw} 为正值;在 Tilt 方向上,人脸向下转动时 θ_{Tilt} 为负值,向上转动时 θ_{Tilt} 为正值。在 Roll 方向上,眼睛沿线与 X 轴正方向夹角在第 1 象限 θ_{Roll} 为正,眼睛沿线与 X 轴正方向夹角在第 4 象限 θ_{Roll} 为负。

首先对仅在 Yaw 方向有转动的情况利用 BP 网络进行训练。 θ_{Yaw} 是 Yaw 方向转动角度,单位为度; $\ln(m)$ 是由手工定位特征点算得的面积比 m ; T 是教师指导信号,如表 1 所示。

表 1 绕 Yaw 方向单一转动训练样本

Tab. 1 Yaw-turn training samples

θ_{Yaw}	$\ln(m)$	$T(\theta_{\text{Yaw}}/50)$
10	1.103 4	0.2
20	1.224 7	0.4
30	1.514 3	0.6

部分 Yaw 方向旋转的情况测试结果如表 2 所示。其中 Out 为神经网络输出, $Y_{\text{aw_estimate}}$ 的平均误差为 1.67° 。

表 2 绕 Yaw 方向单一转动估计结果

Tab. 2 Yaw-turn estimate results

θ_{Yaw}	$\ln(m)$	Out	$Y_{\text{aw_estimate}}$
-5	0.972 2	-0.086	-4.3
15	1.125 8	0.292	14.6
25	1.424 4	0.534	26.7

用同样的方法,对仅有在 Tilt 方向有转动的情况利用 BP 网络进行训练。 $T_{\text{ilt_estimate}}$ 的平均误差为 1.53° 。

最后,对 Yaw 方向和 Tilt 方向复合转动的情况利用 BP 网络训练。部分训练样本如表 3 所示。

表 3 部分复合转动训练样本

Tab. 3 Multiple-turn training samples

θ_{Yaw}	θ_{Tilt}	$\ln(m)$	$\ln(n)$	$T(\theta_{\text{Yaw}}/50)$	$T(\theta_{\text{Tilt}}/50)$
0	20	1	0.855 7	0	0.4
10	0	1.124 8	1	0.2	0
10	20	1.198 7	0.841 9	0.2	0.4
20	5	1.302 4	0	0.4	0.1

部分复合转动测试结果如表 4 所示。

表 4 部分复合转动估计结果

Tab. 4 Multiple-turn estimate results

θ_{Yaw}	θ_{Tilt}	$\ln(m)$	$\ln(n)$	$Y_{\text{aw_estimate}}$	$T_{\text{ilt_estimate}}$
10	30	1.116	0.656	8.8	32.6
20	20	1.29	0.835	17.5	19.8
30	15	1.54	0.926	34.2	15.8

对 Yaw 方向 $-50^\circ \sim +50^\circ$ 范围内、Tilt 方向 $-50^\circ \sim +50^\circ$ 姿态范围内的图像做了实验,估计复合转动角度时,平均误差 $Error_{\text{Yaw}} = 3.8^\circ$, $Error_{\text{Tilt}} = 3.1^\circ$ 。

部分实验结果如图 5 所示。

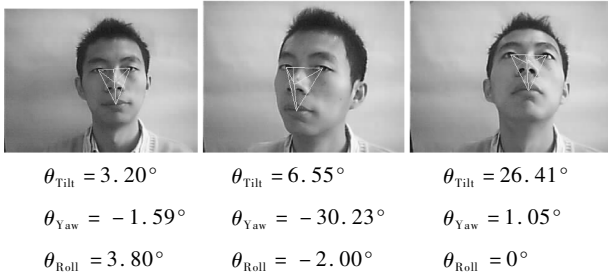


图 5 部分实验结果

Fig.5 Experimental results

从实验结果可以看出,该方法依据单幅图片,不依赖于摄像机焦距和人脸离摄像机距离远近的影响,而且在特征点的选取方面,只利用了 4 个特征点就能方便地定位姿态。

3 原型系统实现

在 Visual C++ 及 OpenGL 环境下开发基于人脸姿态的虚拟环境交互漫游系统原型,该系统处理流程如图 6 所示。

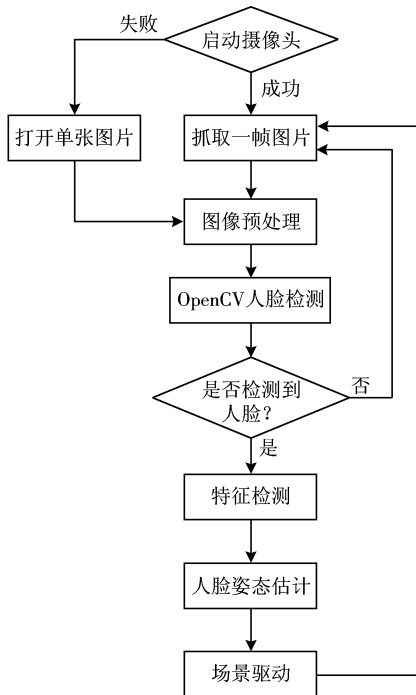


图 6 原型系统处理流程图

Fig.6 Processing flow chart of prototype system

系统预定义了 4 种姿态作为识别对象,每种姿态对应不同的含义,分别是向上、向下、向左、向右。

若在摄像头前作出“向左”的姿态,则场景中视点向左转动;若为“向右”姿态,场景中视点则向右转动;其余姿态也会使视点作相应的转动。图 7 是接受“向左”姿态命令时系统响应的结果。



(a) 识别前场景



(b) “向左”姿态命令时响应的结果

图 7 “向左”姿态识别结果

Fig.7 Recognition result of left

系统原型在 P4 2.66GHz,512MB 内存的机器上加以实现,系统能及时对姿态的变化作出反应,处理速度达到 10 帧/s,基本满足实时性的要求。

4 结 论

本文讨论了人脸姿态估计的方法,根据估计结果驱动虚拟场景运动。详细阐述了基于人脸姿态的虚拟环境交互实现方法,并在 Visual C++ 和 OpenGL 环境中加以实现。下一步工作中,解决将角度的估计值与场景中的视点方位精确对应,以及大角度人脸偏转的识别问题。同时,与语音、手势相结合形成多通道用户交互系统,应用于 3D 游戏,教育娱乐软件中。

参考文献 (References)

- [1] Sebe N, Lew M S, Huang T S. The state-of-the-art in human-computer interaction [J]. *Lecture Notes in Computer Science*, 2004, 3058: 1-6.
- [2] Ferrario V F, Sforza C, Serrao G, et al. Active range of motion of the head and cervical spine: a three-dimensional investigation in healthy young adults [J]. *Journal of Orthopaedic Research*, 2002, 20(1): 122-129.
- [3] Liang Guoyuan, Cha Hongbing, Liu Hong. Face pose estimation based on 3D models and affine correspondences [J]. *Chinese Journal of Computers*, 2005, 28(5): 792-800. [梁国远, 查红彬, 刘宏. 基于三维模型和仿射对应原理的人脸姿态估计方法 [J]. *计算机学报*, 2005, 28(5): 792-800.]
- [4] Hu Bufa, Qiu Limei. 3D Face pose estimation based on multi-points model [J]. *Journal of Image and Graphics*, 2008, 13(7): 1353-1358. [胡步发, 邱丽梅. 基于多点模型的 3D 人脸姿态估计方法 [J]. *中国图象图形学报*, 2008, 13(7): 1353-1358.]
- [5] Athanasios Nikolaidis, Ioannis Pitas. Facial feature extraction and determination of pose [J]. *Pattern Recognition*, 2000, 33(11): 1783-1791.
- [6] Ballard P, Stockman G C. Controlling a computer via facial aspect [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1996, 25(4): 670-678.
- [7] Ho S Y, Huang H L. An analytic solution for the pose determination of human faces from a monocular image [J]. *Pattern Recognition Letters*, 1998, 19(11): 1045-1054.
- [8] Horprasert A T, Yacoob Y, Davis L S. Computing 3-D head orientation from a monocular image sequence [C] // *Proceedings of 2th International Conference on Automatic Face and Gesture Recognition*. Vermont, USA: IEEE Computer Society, 1996: 242-247.
- [9] Yilmaz A, Shafique K, Shah M. Estimation of rigid and non-rigid facial motion using anatomical face model [C] // *Proceedings of 16th International Conference on Pattern Recognition*. Washington, USA: IEEE Computer Society, 2002: 1-4.
- [10] Rees D, Hogg T, Talhami H. Three-dimensional pose from two-dimensional images: a novel approach using synergetic networks [C] // *Proceedings of IEEE International Conference on Neural Networks*. Washington, USA: IEEE Computer Society, 1995: 1140-1144.
- [11] Wei Y, Fradet L, Tan T. Head pose estimation using gabor eigenspace modeling [C] // *Proceedings of the IEEE International Conference on Image Processing*. Washington, USA: IEEE Computer Society, 2002: 281-284.
- [12] Zhang Hongmei. Research on technology of Face Detection and Facial Feature Localization [D]. Chongqing: Chongqing University, 2005. [张红梅. 人脸检测及面部特征定位技术研究 [D]. 重庆: 重庆大学, 2005.]
- [13] Li Song, Ping Xijian, Ding Yihong. Application of open source computer vision library [J]. *Computer Application and Software*, 2005, 22(8): 134-136. [黎松, 平西建, 丁益洪. 开放源代码的计算机视觉类库 OpenCV 的应用 [J]. *计算机应用与软件*, 2005, 22(8): 134-136.]
- [14] Zhang Honglin. VC Digital Image Pattern Recognition Technology and Engineering Practice [M]. Beijing: Posts & Telecom Press, 2003. [张宏林. *Visual C++ 数字图像模式识别及工程实践* [M]. 北京: 人民邮电出版社, 2003.]
- [15] Hagan M T, Menhaj M B. Training feedforward networks with the marquardt algorithm [J]. *IEEE Transactions on Neural Networks*, 1994, 5(6): 989-993.