第 28 卷 第 6 期
2005 年 6 月
计 算 机 学 报
CHINESE JOURNAL OF COMPUTERS
Vol. 28 No. 6
June 2005

基于 SFM 算法的三维人脸模型重建

王 琨 郑南宁
(西安交通大学人工智能与机器人研究所 西安 710049)

摘 要 提出了一种根据两幅正面人脸图像和一幅侧面图像重建人脸三维模型的算法. 该算法主要包括 4 个步骤:寻找匹配点;采用 SFM 算法计算出特征点的三维坐标,并组成稀疏的三维网格结构;采用分步紧支撑径向基函数进行三维插值,得到三维模型;最后根据多分辨图像拼接算法生成纹理图像并将其映射到三维模型上,从而增强真实感. 与其它算法相比,该算法最大的不同之处在于匹配点的寻找. 匹配点的准确与否直接影响 SFM 算法结果的正确性. 许多寻找匹配点的算法如角点匹配算法,在处理人脸图像时得到的结果并不稳定. 这是因为人脸图像上包含了许多低纹理和重复纹理区域. 大多数算法将代表人脸结构基本特征的基准模型运用在重建过程的最后一步,通过三维逼近运算,得到最终的重建模型. 而该算法将反映人脸共性特征的几何对称性和规律性运用到匹配点的寻找中,能够快速准确地找出 SFM 算法需要的匹配点. 用户使用普通照相机拍摄到的图像经本算法的处理后就可以得到相应的三维人脸结构.

关键词 三维重建;SFM 算法;三维视觉;径向基函数;图像拼接
中图法分类号 TP391

3D Face Modeling Based on SFM Algorithm

WANG Kun ZHENG Nan-Ning
(*Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an* 710049)

Abstract This paper proposes a whole scheme to construct textured 3D face models from two views with a few user interactions. As the human face contains many low texture and homogeneous areas, some algorithms such as corner matching are unstable and may fail sometimes. This paper uses the face definition parameters and the symmetry of human face as prior knowledge to find reliable correspondences between two pictures, while most SFM algorithms use the generic model as a modulator in the post-processing steps. According to the correspondences, a multistage SFM approach is used to reconstruct the structure. Then the authors use the RBFCS algorithm to interpolate more 3D points according to the scattered feature points. Multistage image mosaic technique is used to get the texture. A user with an ordinary camera can use the algorithm to generate his realistic face model in a personal computer.

Keywords 3D reconstruction; structure from motion algorithm; 3D vision; radial based function; image mosaic

1 引 言

人脸的三维模型能够广泛应用于计算机游戏、电影及视频通信等不同领域. 真实感人脸模型的构造是计算机视觉研究的一个重要问题. 人脸不仅具有极其复杂的几何结构而且脸部包含了许多细小的纹理和褶皱,因此纹理结构也十分复杂.

收稿日期:2004-06-09;修改稿收到日期:2005-04-03. 本课题得到国家"八六三"高技术研究发展计划项目基金(2001AA114202)和国家自然科学基金(60205001,60021302)资助. 王 琨,男,博士研究生,主要研究方向为图像处理. E-mail: wkques@sina.com. 郑南宁,男,1952 年生,教授,博士生导师,研究方向为人工智能、机器人和图像处理.

根据数据的来源可以将人脸重建算法大致分为以下几类：目前应用最为广泛的是采用激光扫描仪，如 Cyber ware Color Digitizer 进行扫描，并通过相应算法（如文献[1,2]）进行处理，得到重建模型. 但是三维激光扫描仪的价格十分昂贵. Garcia 采用结构光的方法[3]进行三维重建. 文献[4,5]采用立体对应算法从二维图像中提取出人脸的三维结构信息. 由于人脸部包含大量的低纹理和相似纹理区域，要得到准确的深度图是十分困难的，而且校准摄像机拍摄立体对应图像也是一项复杂的工作.

根据图像序列进行重建的方法成为近年的研究热点[6~9]. Pighin 等人[7]提出的方法是通过用户在多幅图像中标定特征点，从而重建人脸结构的算法. LEE 等[10]提出了由两幅正交人脸图像估计几何结构的算法. Zhang[11]等提出的方案非常有意义，同时启发了本文的工作. 他们的算法是：首先提示用户在拍摄到的两幅人脸图像中各选取 5 个特征点；然后通过角点匹配的算法，寻找出更多的匹配特征点；通过 SFM 算法计算出这些特征点的三维坐标，并重建三维模型；最后将该模型与基准人脸模型进行三维逼近，得到最终结果. 这一方案比较成功，但有时也会失效. 原因是角点匹配算法失效时引起整个重建工作的失败. 角点匹配算法用在人脸图像上是不稳定的，包括错误匹配和寻找到的角点数量不稳定两个方面. 而 SFM 算法对错误的匹配点比较敏感. 因此对 SFM 算法来说寻找可靠的匹配点是一项十分关键的步骤.

人脸的结构具有很强的对称性. MPEG4 中的 FDP 对人脸的基本性质进行了详细的定义. 本文就是利用这些对称性和规律性来寻找出可靠的匹配点.

本文提出了一套完整的人脸模型重建方案，根据这一方案只需普通的照相机和一台计算机就可以重建出人脸的三维结构：首先通过照相机拍摄人脸的两幅正面图像，并提示用户标定特征点；然后根据 FDP 的定义和人脸的对称性推导出更多的匹配点. 大多数的人脸重建算法都是先构造出一个粗略的三维结构，然后将人脸结构具有的共性特征作为一系列约束来校正前面的模型，例如逼近基准模型的算法. 本文的思路是将人脸结构具有的共性特征作为先验信息运用到匹配点的寻找中，从而得到可靠的匹配点. 与文献[8]相比，本算法采用的匹配点的数量少，但准确度较高，而且这些特征点包含了个性人脸结构的特征信息.

2　系统组成

图 1 为本算法的框图. 本算法主要包括以下几个步骤：第一步是数据采集. 首先拍摄两幅人脸的正面图像，并提示用户在这两幅图像中标定一些特征点；然后再拍摄一幅侧面图像，用于生成纹理图像；第二步是根据已知的特征点，运用 FDP 中的定义和人脸结构的规律性推导出更多的匹配点；第三步，采用 SFM 算法估计两幅图像间的旋转矩阵 $\boldsymbol{R}$ 和位移矩阵 $\boldsymbol{t}$，并计算出特征点的三维坐标. 与文献[1,8]中的网格结构相比，本算法采用的匹配点数量较少，只能组成稀疏的网格结构. 因此在第四步中采用基于分步紧支撑径向基函数（RBFCS）[12~14]的插值算法，对这些稀疏的三维特征点进行插值. 第五步，生成纹理图像. 本文采用多分辨图像拼接算法[15]，从正交的两幅人脸图像生成纹理图像，并将其映射到三维模型上，得到具有真实感的三维人脸模型.

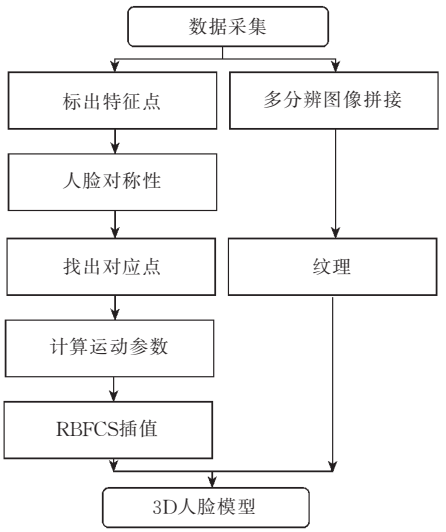


图 1　系统框图

3　算　法

3.1　匹配点

首先拍摄一张人脸的正面图像，然后将头向上抬起拍摄第二张图像，如图 2. 这样做的原因是：(1)将头向上抬起是一个自然并且容易完成的动作；(2)当头部抬起后，所有三维点的 x 坐标（横向，如图 2）保持不变. 本算法中将利用这一性质来推导匹配点；(3)由于头部在几何结构上是左右对称的. 因此本算法只重建半边模型，另一半通过对称关系获得；

而左右两侧人脸的差异将通过纹理的不同来体现.

接着提示用户在图像上标定出 17 个特征点,包括内眼角、鼻尖、嘴角等. 并将这些点标记为 $P=\{p_i \mid i=1,2,\cdots,N\}$,其中 $N=17$. 图 2 中较大的白色圆点是由用户标定的特征点. MPEG4 中的 FDP 对人脸的基本参数进行了定义. 表 1 是 FDP 的部分内容. 下面以右眼为例,说明本算法寻找匹配点的过程.

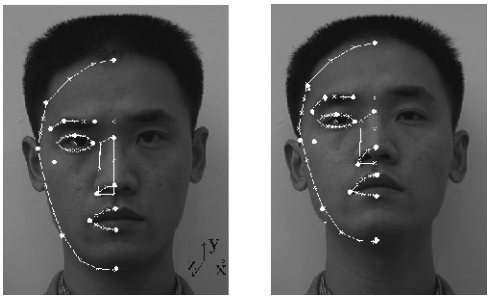


图 2 两幅正面图像

表 1 部分 FDP 参数

特征点		建议坐标约束		
	描述	x	y	z
4.1	左眉的右角点			
4.2	右眉的左角点			
4.3	左眉最高点	$(4.1.x+4.5.x)/2$ 或左眉轮廓最高点的 x 坐标		
4.4	右眉最高点	$(4.2.x+4.6.x)/2$ 或右眉轮廓最高点的 x 坐标		

1. 通过与用户的交互,可以得到 P_1 和 P_2 点的坐标. 根据表 1 和眼睛的对称性可以求出另外两个顶点以及中心点的坐标. 然后通过构造一个椭圆方程来近似描述右眼上的点;

2. 在椭圆上再找出 12 个点,这些点的 x 坐标是 P_2 和 P_4 之间的等分点,并记下这些点的坐标. 如图 3(a). 用这 17 个点来代表右眼的位置;

3. 对图 3(b)进行同样的操作,可得 P_1' 和 P_2' 等 17 个点的坐标;

4. 由于两幅图像中特征点的 x 坐标没有改变,而且根据用户的标定可知(P_1,P_1') 和(P_2,P_2')是匹配点,因此其它的点也相应的是匹配点.

(a) (b)

图 3 眼部特征点的匹配

3.2 SFM 算法

在本文中,将向量 x 的齐次坐标记为 $\tilde{x}$. 三维点 $\boldsymbol{M}=[x,y,z]^{\mathrm{T}}$ 的齐次坐标记为 $\widetilde{\boldsymbol{M}}=(x,y,z,1)^{\mathrm{T}}$,其对应的图像点 $\boldsymbol{m}=[u,v]^{\mathrm{T}}$ 的齐次坐标记为 $\tilde{\boldsymbol{m}}=[u,v,1]^{\mathrm{T}}$. 它们之间的关系满足公式(1).

$$\lambda\tilde{\boldsymbol{m}}=\boldsymbol{A}\boldsymbol{V}\boldsymbol{\Omega}\widetilde{\boldsymbol{M}} \tag{1}$$

其中 $\boldsymbol{A}=\begin{pmatrix}\alpha & \nu & u_0\\ 0 & \beta & v_0\\ 0 & 0 & 1\end{pmatrix}$, $\boldsymbol{V}=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\end{pmatrix}$, $\boldsymbol{\Omega}=\begin{pmatrix}R & t\\ 0^{\mathrm{T}} & 1\end{pmatrix}$, λ 是一个常量.

矩阵 $\boldsymbol{A}$ 中的元素是摄像机的内部参数,通过标定摄像机获得[16]. 矩阵 $\boldsymbol{\Omega}$ 是三维坐标变换矩阵,包括旋转矩阵 $\boldsymbol{R}$ 和位移矩阵 $\boldsymbol{t}$. 定义基本矩阵 $\boldsymbol{E}$ 为 $\boldsymbol{E}=[t]\times\boldsymbol{R}$.

根据 3.1 节的算法可以求出匹配点集合. 因此三维重建问题就转化为已知匹配点的 SFM 问题. 本算法采用以下步骤来求解基本矩阵 $\boldsymbol{E}$、旋转矩阵 $\boldsymbol{R}$ 和位移矩阵 $\boldsymbol{t}$[11].

1. 用 8 点算法估计基本矩阵 $\boldsymbol{E}$ 的初值,记为 $\boldsymbol{E}_1$;

2. 应用 zero-determinant 约束条件,由 $\boldsymbol{E}_1$ 推出 $\boldsymbol{E}_2$;

3. 求解使特征点和极线(epipolar lines)间的距离最小的基本矩阵,记为 $\boldsymbol{E}_3$;

4. 从基本矩阵 $\boldsymbol{E}_3$ 中估计旋转矩阵 $\boldsymbol{R}$ 和位移矩阵 $\boldsymbol{t}$. 因为 $\boldsymbol{E}^{\mathrm{T}}\boldsymbol{t}=0$,因此 $\boldsymbol{t}$ 如下求解:

求解 $\min\limits_{t}\|\boldsymbol{E}^{\mathrm{T}}\boldsymbol{t}\|^2$, 其中 $\|\boldsymbol{t}\|=1$,

基本矩阵的定义[11,17]为 $\boldsymbol{E}=[\boldsymbol{t}]\times\boldsymbol{R}$,因此 $\boldsymbol{R}$ 如下求解:

$\min\limits_{R}\|\boldsymbol{E}-[\boldsymbol{t}]\times\boldsymbol{R}\|^2$,其中 $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}=\boldsymbol{I}$ 且 $\det(\boldsymbol{R})=1$.

5. 通过最大相似度估计方法来进一步估计运动和结构参数;

$$\hat{\theta}=\arg\max_{\theta}\log(P(X\mid\theta)),$$

其中 $\hat{\theta}$ 是运动和结构参数 θ 的最大相似度估计,X 是匹配点的集合;

6. 已知$(\boldsymbol{R},\boldsymbol{t})$后可以计算出特征点在三维空间中的坐标值 P. 假设(m,m')是 P 点在两幅图像上的投影,通过公式(2)校正 P 点的坐标值.

$$P=\arg\min_{p}(\|m-\hat{m}\|^2+\|m'-\hat{m}'\|^2) \tag{2}$$

第 1 步是一个最小平方问题;第 3,5,6 步是非线性最小化问题,可以通过 Levenberg-Marquardt 算法求解;第 4 步可以采用 quaternion representation[11] 算法求解. 图 4 是由特征点的 3D 坐标组成的网格结构.

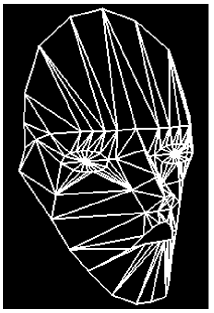


图 4　稀疏的网格结构

3.3　三维数据插值

由图 4 可以看出，这是一个稀疏的三维网格结构. 在本节中将根据这些特征点插入更多的三维点[12,14].

假定特征点为 $X=\{x_1,x_2,\cdots,x_N\}\in R^3$. 将问题定义为寻找一个插值函数 $F:R^3\to R^3$. 定义

$$F(x):=\sum_{j=1}^{n}C_j\cdot\phi(x-x_j),$$

其中 $C_j\in R^3$ 为第 j 个基函数的权值. 基函数 $\phi:R^3\to R$ 在 R^3 空间上是正定的，且 $\phi(x)=\phi(\|x\|)=\phi(r)$. 假设 $F(x_i)=y_i$ 为已知的对应点. 则

$$\begin{cases}c_1\phi(\|x_1-x_1\|)+c_2\phi(\|x_1-x_2\|)+\cdots+\\ \quad c_N\phi(\|x_1-x_N\|)=y_1\\ \cdots\\ c_1\phi(\|x_N-x_1\|)+c_2\phi(\|x_N-x_2\|)+\cdots+\\ \quad c_N\phi(\|x_N-x_N\|)=y_N\end{cases}$$

记为

$$\boldsymbol{A}_{x,\phi}\boldsymbol{C}=\boldsymbol{Y}\tag{3}$$

其中 $\boldsymbol{A}=\begin{bmatrix}\phi(\|x_1-x_1\|) & \cdots & \phi(\|x_1-x_N\|)\\ & \vdots & \\ \phi(\|x_N-x_1\|) & \cdots & \phi(\|x_N-x_N\|)\end{bmatrix}$，$\boldsymbol{C}=(c_1,c_2,\cdots,c_N)^{\mathrm{T}}$，$\boldsymbol{Y}=(y_1,y_2,\cdots,y_N)^{\mathrm{T}}$.

假如 $\boldsymbol{A}$ 正定，则方程(3)有解，$\boldsymbol{U}=\boldsymbol{A}^{-1}\boldsymbol{V}$.

许多文章都对基函数的构造进行了研究[12~14,18]. 其中一种称为紧支撑径向基函数，不仅具有计算量小而且能够较好地表现出函数的局部特性. 式(4)中给出了一类经典的基函数[14].

$$\begin{cases}\Phi_{3,0}=(1-r)^7(5+35r+105r^2+147r^3+\\ \qquad 101r^4+35r^5+5r^6)\in C^6\\ \Phi_{3,1}=(1-r)^6(6+36r+82r^2+72r^3+\\ \qquad 30r^4+5r^5)\in C^4\\ \Phi_{3,2}=(1-r)^5(8+40r+48r^2+25r^3+\\ \qquad 5r^4)\in C^2\\ \Phi_{3,3}=(1-r)^4(16+29r+20r^2+5r^3)\in C^0\end{cases}\tag{4}$$

在人脸结构中，特征点的分布是不均匀的，眼睛和嘴部多而脸颊部分的特征点少. 在本算法中将这些特征点分层，即分成一系列具有包含关系的集合.

$$S_0\subset S_1\subset\cdots\subset S_t=S.$$

将插值算法分部进行：在最小点集即最稀疏的点集上使用平滑性好的基函数插值，并且使用较大的支持半径；当点集稠密时，使用不太平滑的基函数，并减小支持半径. 而最终的插值函数为每一步所得插值函数的和，$F:=\sum_{k=1}^{t}F_k$.

图 5 为插值后得到的较密的网格结构.

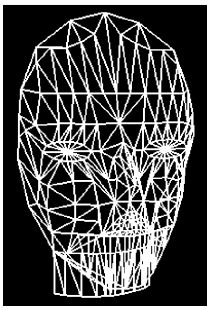


图 5　插值后的网格结构

3.4　纹理生成

本文采用多分辨图像拼接算法[10,15]从正面和侧面两幅图像生成脸部的纹理. 多分辨图像拼接算法中主要包括两种运算“REDUCE”和“EXPAND”.

在“REDUCE”过程中使用低通滤波器将图像分解为一系列子图像.

假定 G_0 是初始图像，则

$$G_l=\mathrm{REDUCE}(G_{l-1})\ (0<l<N),$$

并定义

$$G_l(i,j)=\sum_{m=1}^{5}\sum_{n=1}^{5}w(m,n)\cdot G_{l-1}(2i+m,2j+n),$$

其中权值 $w(m,n)$ 满足

$$\begin{cases}w(m,n)=\hat{w}(m)\hat{w}(n)\\ \hat{w}(0)=a;\ \hat{w}(-1)=\hat{w}(1)=b;\\ \quad\hat{w}(-2)=\hat{w}(2)=c\\ a+2b+2c=1\\ a+2c=2b\end{cases}\tag{5}$$

“EXPAND”运算完成与“REDUCE”运算相反的功能. 令 $G_{l,k}$ 表示将 G_l 扩展 k 次后得到的结果，则 $G_{l,0}=G_l$；

当 $k>0$ 时，$G_{l,k}=\mathrm{EXPAND}[G_{l,k-1}]$，

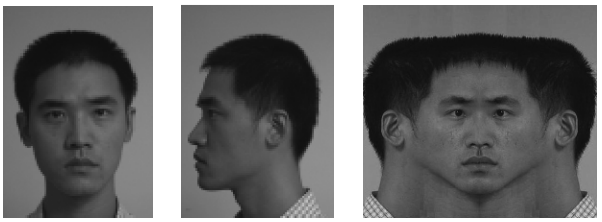
$$G_{l,k}(i,j)=4\sum_{m=2}^{2}\sum_{n=-2}^{2}G_{l,k-1}((2i+m)/2,(2j+n)/2)\tag{6}$$

在式(6)中,仅考虑当$(2i+m)/2$和$(2j+n)/2$为整数时的项.由此可以推出一系列子图像L_0,L_1,…,L_N.

当$0<l<N$时,

$$L_l = G_l - \mathrm{EXPAND}[G_{l+1}] = G_l - G_{l+1,1} \quad (7)$$

然后在每一级上拼接正面和侧面两幅图像,并进行"EXPAND"运算,最终得到纹理图像,如图6(c).



(a)正面图像 (b)侧面图像 (c)脸部纹理

图6 纹理生成

4 实验结果及结论

我们采用这一算法进行了人脸模型的重建.选择图2所示的两幅正面图像,用户在这两幅图像上标定特征点.本算法根据FDP中的参数和人脸的对称性及规律性,推导出更多的匹配点,如图3所示.图4是重建的稀疏网格结构.图5是经过插值后的结构.图6(a)(b)是用来生成纹理图像的正面和侧面图像.图6(c)是最终的纹理图像.所有的图像都是用普通的照相机在一般环境下拍摄的.在图7中,将重建模型与实际拍摄到的图像在不同角度进行了比较.

图7 重建模型与实际照片的比较

本算法得到的模型具有较好的正面和侧面结构.由于本算法采用两幅正面图像来重建人脸结构,因此缺少后半面的结构以及耳朵、牙齿等部分的结构.在实验中参考了Keith Waters的Geoface.

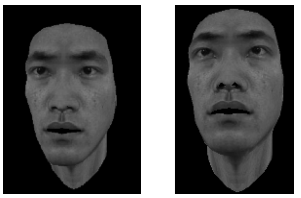


图8 说话时的模型

参 考 文 献

1 Lee Yuen-Cheng, Terzopoulos D., Walters K.. Realistic modeling for facial animation. In: Proceedings of the SIGGRAPH95, Los Angeles, 1995, 55～62

2 Lee Y. C., Terzopoulos D., Waters K.. Constructing physics-based facial models of individuals. In: Proceedings of Graphics Interface, Toronto, 1993, 1～8

3 Garcia E., Dugelay J. L.. Low cost 3D face acquisition and modeling. In: Proceedings of the International Conference on Coding and Computing, Las Vegas, 2001, 657～661

4 Lengagne R., Fua P., Monga O.. 3D stereo reconstruction of human faces driven by differential constraints. Image and Vision Computing, 2000, 18(4): 337～343

5 Lengagne R., Tarel J. P., Monga O.. From 2D images to 3d face geometry. In: Proceedings of IEEE Second International Conference on Automatic Face and Gesture Recognition, Kiuington, VT, 1996, 301～306

6 Roy Chowdhury A., Chellappa R., Krishnamurthy S., Vo T.. 3D face reconstruction from video using a generic model. In: Proceedings of IEEE International Conference on Multimedia and Expo 2002, Lausanne, Switzerland, 2002, 1: 26～29

7 Pighin F., Hecker J., Lischinskiy D., Szeliski R., Salesin D. H.. Synthesizing realistic facial expressions from photographs. In: Proceedings of the SIGGRAPH98, Orlando, FL, USA, 1998, 75～84

8 Liu Zi-Cheng, Zhang Zheng-You *et al*.. Rapid modeling of animated faces from video. Microsoft Technical Report MSR-TR-2000-11, 2000

9 Ambrosini L., Costa M., Lavagetto F., Pockaj R.. 3D head model calibration based on MPEG-4 Parameters. In: Proceedings of the 6th IEEE International Workshop on Intelligent Signal Processing and Communication Systems, Melbourne, 1998, 626～630

10 LEE Won-Sook, Nadia Magnenat-Thalmann. Fast head modeling for animation. Image and Vision Computing, 2000, 18(18): 355～364

11 Zhang Zheng-You. A new multistage approach to motion and structure estimation by gradually enforcing geometric constraints. In: Proceedings of the ACCV'98, Hong Kong, 1998, 567～574

12 Schaback R.. Creating surfaces from scattered data using radial basis functions. Mathematical Methods in Computer Aided Geometric Design III, San Diego, CA, USA: Academic Press Inc., 1995

13 Floater M. S., Iske A.. Multistep scattered data interpolation using compactly supported radial basis functions. Journal of Computational and Applied Mathematics, 1996, 73(5): 65～78

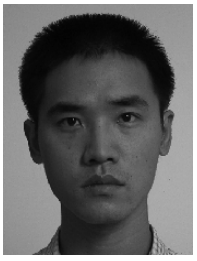
14 Lavagetto F., Pockaj R., Costa M.. Smooth surface interpolation and texture adaptation for MPEG-4 compliant calibration of 3D head models. Image and Vision Computing, 2000, 18(4): 345～354

15 Burt P. J., Adelson E. H.. A multiresolution spline with application to image mosaics. ACM Transaction on Graphics, 1983, 2(4): 217～236

16 Zhang Z.. Flexible camera calibration by viewing a plane from unknown orientations. In: Proceedings of the ICCV 1999, Kerkyra, Corfu, Greece, 1999, 666～673

17 Longuet-Higgins H.. A computer algorithm for reconstructing a scene from two projections. Nature, 1981, (293): 133～135

18 Saligon O., Le Mehaute A., Roux C.. Facial expression simulation with RBF. In: Proceedings of International Workshop on SNHC and 3D Imaging, Rhodes, Greece, 1997, 233～236



WANG Kun, born in 1973, Ph. D. candidate. His research interests include image processing and virtual reality.

ZHENG Nan-Ning, born in 1952, professor, Ph. D. supervisor. His research interests include artificial intelligence, robotics and image processing etc.

Background

This paper is part of the research about virtual reality platform. The research was supported by the National High Technology Research and Development Program (863 Program) under grant No. 2001AA114202 and National Natural Science Foundation of China under grant No. 60205001 and 60024301. It proposes a whole scheme to construct textured 3D face models from two views with a few user interactions. A user with an ordinary camera can use the algorithm to generate his realistic face model in a personal computer.