

Research on 3D Reconstruction of Face Based on Binocular Stereo Vision and Deformation Model

Xuhao Xiao¹, Junjie Li¹, Haihui Wang^{1,2*}, Qing Hou¹, Zhi Ruan¹

¹School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan Hubei

²Hubei Provincial Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan Hubei

Email: *wanghaihui69@sina.com

Received: Jun. 28th, 2018; accepted: Jul. 12th, 2018; published: Jul. 20th, 2018

Abstract

People usually perceive things from a three-dimensional perspective, requiring the computer to automatically reconstruct the corresponding three-dimensional shape model using two-dimensional human face images. The three-dimensional face reconstruction based on 3DMM under binocular vision is to use SIFT algorithm for feature point matching and disparity calculation to obtain three-dimensional coordinates. The three-dimensional coordinates are applied directly to the face deformation model, which overcomes the traditional method of finding two-dimensional and three-dimensional coordinates, convert the relationship, the disadvantage of low precision, thus reconstruct a three-dimensional face model with a strong sense of reality.

Keywords

Face Detection, Three-Dimensional Reconstruction, SIFT Algorithm, Deformation Model

基于双目立体视觉和形变模型的人脸三维重建研究

肖栩豪¹, 李俊杰¹, 王海晖^{1,2*}, 侯青¹, 阮志¹

¹武汉工程大学计算机科学与工程学院, 湖北 武汉

²武汉工程大学智能机器人湖北省重点实验室, 湖北 武汉

Email: *wanghaihui69@sina.com

收稿日期: 2018年6月28日; 录用日期: 2018年7月12日; 发布日期: 2018年7月20日

*通讯作者。

文章引用: 肖栩豪, 李俊杰, 王海晖, 侯青, 阮志. 基于双目立体视觉和形变模型的人脸三维重建研究[J]. 图像与信号处理, 2018, 7(3): 161-169. DOI: 10.12677/jisp.2018.73019

摘要

人类通常对事物的感知是从三维的角度出发,要求计算机也能够利用二维人脸图像自动重建相应的三维形状模型。双目下基于形变模型的人脸三维重建,是通过SIFT算法进行特征点匹配、视差计算得到三维坐标,将三维坐标直接应用于人脸形变模型,克服了传统方法寻找二维与三维坐标的大致转换关系,精度低的缺点,从而重建出具有较强真实感的三维人脸模型。

关键词

人脸检测, 三维重建, SIFT算法, 形变模型

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

人类通常对事物的感知是从三维的角度出发的,而且由于人类特殊的视觉系统,即使是二维的人脸照片,人类也能自动生成其三维形状。如果计算机可以模拟实现和人类视觉系统相同的功能,那么计算机对人脸图像的认知、检索和分辨能力都将有很大的提高。而这这就要求计算机能够利用二维人脸图像自动重建相应的三维形状模型[1]。

人脸三维重建技术改变了传统的人脸图像检索和识别手段,可以根据实际需求模拟出各种情况下的人脸状态。而且三维模型比普通的二维图像更接近真实状态,能够更好地模拟实际情况,更便于理解和操作。迄今为止,在国内和国外仍然有很多研究机构都在进行三维人脸建模的研究工作,目前人脸三维模型已经广泛应用在医学手术、三维动画制作、人脸识别、视频通话等领域[2]。人脸三维模型的重建已成为计算机视觉的一个重要研究领域和挑战性课题。关于人脸三维重建的方法较多,结合双目立体视觉和形变模型的优势,本文提出了一种实现人脸三维重建的方法。

2. 系统组成

本系统由七部分组成,分别是相机标定功能、图像获取功能、图像处理功能、图像校正功能、人脸鉴别功能、人脸匹配功能和三维重建功能(图 1)。相机标定使用鲁棒性和准确性较高的张正友标定方法;实验使用图像通过硬件工业相机模拟双目进行获取;利用滤波技术将图像进行初步处理,以减少噪音,突出边界等作用,配合标定出的相机内外参数,对图像进行畸变和立体校正;人脸识别使用。

3. 识别算法与重建

3.1. 匹配算法

SIFT 算法是 David Lowe 于 1999 年提出的基于局部区域的特征描述子,能同时解决两幅图像之间发生的各种任意的旋转和平移,以及不同角度的仿射变换等情况下的匹配问题;另外 SIFT 扩展性强,可以很方便的与其他形式的特征向量进行联合[3]。

SIFT 特征匹配分为三个工序:提取关键点,生成特征描述器,特征点匹配。SIFT 算法先求解图像每

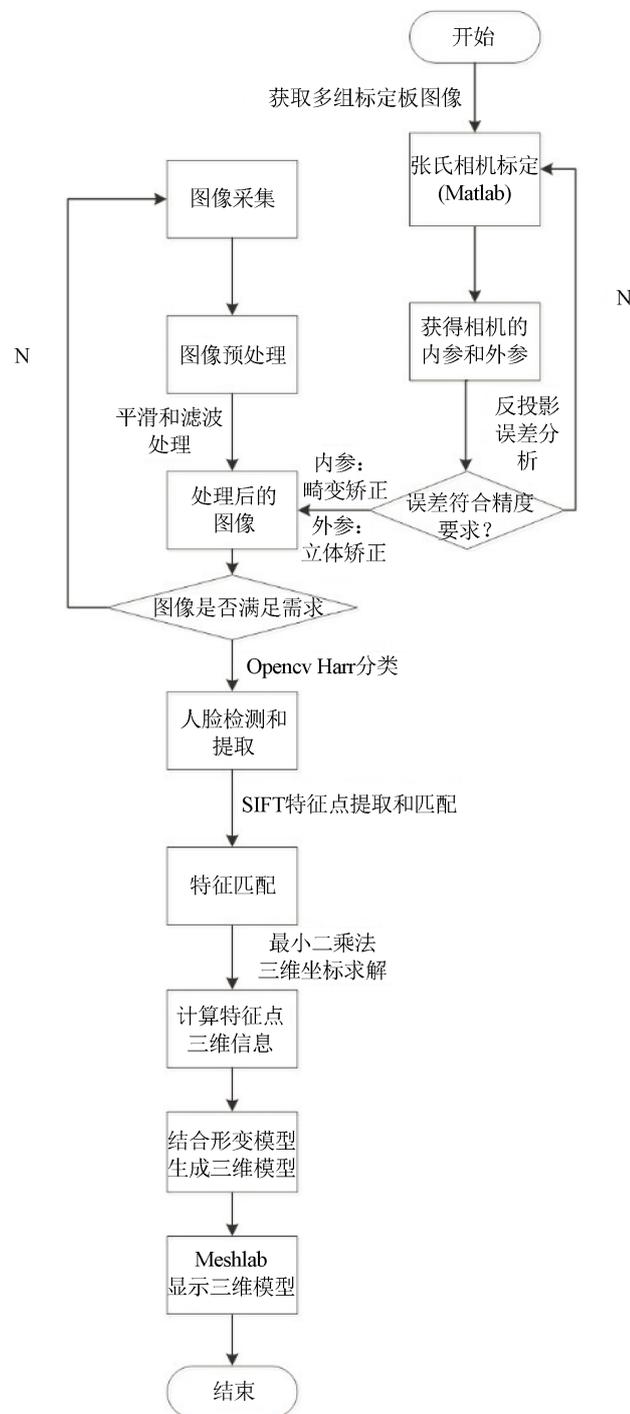


Figure 1. System composition

图 1. 系统组成图

一个像素点的 Hessian 矩阵，并提取图像的特征点，然后进行计算高斯卷积[4]，将 Hessian 矩阵的行列式中的区域极大值点认为图像特征点。

当提取特征点的时候是在特征点邻接矩形区域内进行，利用 Sobel 算子来计算特征点区域水平和垂直方向的梯度，接着公式(1)和公式(2)计算幅值 g 和方向 θ [5]。

$$g = \sqrt{g_x^2 + g_y^2} \quad (1)$$

$$\theta = \arctan \frac{g_y}{g_x} \quad (2)$$

每个像素点都有一个幅值和方向,对于有颜色的图像会在 R、G、B 三个基色分量上计算梯度,并最终选取最大的幅值作为该像素点的最终幅值。然后对 20×20 (单位为 pixel) 区域划分为 4×4 (或 2×2) 的子区域,每个子域计算 8 bin 直方图(即每 45° 为一个方向块) [6]。最后根据计算出的梯度幅值和方向,将结果根据比例相加在直方图对应的方向块(即对应 bin)中,形成梯度直方图得出该区域相关方向数据。

最后通过特征点描述子之间的欧式距离作为相似性度量结果。首先计算获得模板图的关键点描述子 $R_i = (r_{i1}, r_{i2}, \dots, r_{i128})$ 和实时图的关键点描述子 $S_i = (s_{i1}, s_{i2}, \dots, s_{i128})$ [7], 然后计算任意两描述子相识性度量

$$d(R_i, S_i) = \sqrt{\sum_{j=1}^{128} (r_{ij} - s_{ij})^2} \quad (3)$$

SURF 构造的金字塔图像与 SIFT 有非常大的不同点。SIFT 使用的是 DOG 特征,而 SURF [8] 使用的是 Hessian 矩阵行列式中的近似值图像,使得检测速度变快。图像中某个像素点的 Hessian 矩阵如下:

$$H(f(x, y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \quad (4)$$

对某个像素点都有对应一个 Hessian 矩阵。在计算 Hessian 矩阵前,都需要对其先进行高斯滤波一下 [9]。滤波后再进行求解 Hessian 矩阵,如公式(5)所示:

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \quad (5)$$

是高斯滤波之后的图像在 x 方向的二阶导数,而 $L_{yy}(x, \sigma)$ 、 $L_{xy}(x, \sigma)$ 为普通二阶导数。为了尽可能的减小相应的计算复杂度,就应该使用积分图像,该方法目的是在模糊图像的基础之上把原本的模块再近似。

使用一般方法计算图像的二阶导数比较复杂,所以引入了积分图的方法。它是指当前像素点所在的位置距离原点(0,0)坐标所包围面的所有灰度值之和[10]。计算图像中某一块矩形的区域灰度和 S_x 利用矩形的 4 个顶点 (X_i, Y_i) ($i=1,2,3,4$, 按从上到下、先左后右的顺序)的积分值 $S_{(x,y)}$ 即可。

$$S_x = S(X_1, Y_1) + S(X_4, Y_4) - S(X_2, Y_2) - S(X_3, Y_3) \quad (6)$$

在函数定义之前要计算出每一坐标点原本的积分图像,然后才能计算 Hessian 的特征值情况。因为函数模板是相似的,所以需要修正特征值的计算方式:

$$\det(H_{approx}) = D_{xx} D_{yy} - (0.9 D_{xy})^2 \quad (7)$$

其中 0.9 是一个经验值。

3.2. 三维坐标求解算法

在重建中,实验涉及三个坐标,分别是世界坐标、相机坐标和图像坐标。在重建中,实验涉及三个坐标,分别是世界坐标、相机坐标和图像坐标。世界坐标系是目标物体位置的参考系。在标定时确定标定物的位置,盛放重建后的物体的三维坐标。摄像机坐标系:其是摄像机站在自己角度上衡量的物体的

坐标系。它是与拍摄物体发生联系的桥头堡，世界坐标系下的物体需先经历刚体变化转到摄像机坐标系，然后在和图像坐标系发生关系。它是图像坐标与世界坐标之间发生关系的纽带，沟通了世界上最远的距离。而图像坐标系：其是以摄像机拍摄的二维照片为基准建立的坐标系。用于指定物体在照片中的位置。

空间点的重建不仅与空间点或者空间结构在图像中的位置有关还与摄像机的内部和外部参数有关。三维重建的原理在于三种坐标之间的转换。

实验用最小二乘法来对特征点的三维坐标进行第一次求解。对图像中的特征点的坐标进行处理，使其为齐次坐标。并利用特征算法进行左右图像上的特征点匹配，并记录其齐次坐标，进而计算出投影矩阵，得出特征点的世界坐标。因为实际应用中提取的数据含有噪声，所以采用最小二乘法的方法来再次求解特征点的世界坐标值。

紧接着针对双目立体视觉系统，采用一种特殊的三维空间点求解方法，即基于视差原理的视差测距法两个摄像机平行共面放置，两个摄像机的光轴也互相平行，并且还有一对坐标轴共线，通过使用两个内参数相同的摄像机，使放置的双目视觉系统的基线距离固定，然后再获取摄像机的内参数求解物点的空间三维坐标。

对于图像中的一个特征点，获取其在左右图像的投影点坐标，并且计算出视差。因为两个摄像机中心连线距离即为基线，由公式(8)，计算出特征点的三维坐标。利用第一次测得的特征点坐标与本次测得的坐标进行均值，降低了实验误差，提高了测量精度。

$$\begin{aligned} X_w &= \frac{Z_w}{f} v_1 = \frac{u_1 B}{d} \\ Y_w &= \frac{Z_w}{f} v_1 = \frac{v_1 B}{d} \\ Z_w &= \frac{fB}{d} \end{aligned} \quad (8)$$

其中 (X_w, Y_w, Z_w) 为特征点的三维坐标， (u_1, v_1) 为左图像的投影点坐标， d 为视差， B 为两个摄像机中心连线距离， f 为焦距。

通过采用基于视差测距法和最小二乘法来求解三维坐标，得出实验精度与摄像机焦距、摄像机之间的基线长都有关系，焦距越长，基线越长，精度就越高。验证了三维重建的正确性和可行性，可以有效的应用于计算机视觉的应用中。

3.3. 人脸三维重建算法

人脸检测是人脸识别(Face Recognition)的基础，其经过几十年的发展已渐趋成熟。早期使用人工提取人脸特征，训练分类器，进行人脸检测的方法。缺点是当环境较复杂且变化强烈时，检测效果不理想。后来出现从通用的目标检测中移植过来的人脸检测方法效果较好，并能适应于复杂的环境变化，但很耗时。综合二者的优缺点，出现了基于级联结构的卷积神经网络的人脸检测算法，时间和效率皆优。

由于本实验只需得到人脸的大致区域，对算法精度没有过高要求。重点在于人脸的三维重建。所以本实验是利用 Opencv 提供的接口进行人脸检测并截取人脸。

Opencv 的人脸检测主要通过 Haar 分类器实现，预先训练好物体的检测器(xml 格式)，包括正脸、侧脸、眼睛等特征。大概思路是首先将图像均衡归一化到同样大小，利用多级高检测率低拒绝率的多层树分类器逐步拒绝所有的非人脸部分，最后剩下的就是人脸区域。将人其标记截取出来。

对截取后的人脸利用 SIFT 算法进行特征点提取和匹配，然后利用匹配点坐标计算视差，求出对应点的三维坐标。

实验中提出了一种双目下基于形变模型的人脸重建方法。传统的基于形变模型的三维人脸重建一般分为两部分：一是获取原始人脸数据并建立组合模型；二是针对特定人脸图像使用模型进行匹配实现三维人脸的重建。本实验中的创新之处在于，原始人脸数据是利用三维测量原理计算得到，而非寻找二维与三维坐标的大致转换关系[11]，视差计算得到三维坐标的精度比之高很多。

三维人脸可视为线性子空间，即公式(9)。

$$S_{new} = \sum_{i=1}^m a_i S_i = a_1 S_1 + a_2 S_2 + \cdots + a_m S_m \quad (9)$$

a_i 为组合因子，满足 $\sum_{i=1}^m a_i = 1$ 。

人脸的三维形状信息可用匹配点的坐标表示，即

$$S_i = (x_1, y_1, z_1, x_2, y_2, z_2, \cdots, x_k, y_k, z_k)^T \quad (10)$$

其中 k 为匹配点的数目。直接使用三维人脸数据重建三维人脸，计算量巨大。这里采用的是主成分分析 (PCA) 理论对样本数据库进行降维[12]，得到

$$S_{new} = \bar{S} + P\beta \quad (11)$$

其中 \bar{S} 为平均三维人脸， P 为选取的主成分对应的特征向量集合， β 为形变因子。

采用基于贝叶斯最大后验概率的最优解法，令目标函数为

$$E(\beta) = \|P^f \beta - \Delta d\|^2 + \eta \|\beta\|^2 \quad (12)$$

其中 η 权重系数。根据奇异值分解，有 $P^f = U\Lambda V^T$ ，其中 Λ 为特征值构成的对角矩阵， U 由 $P^f (P^f)^T$ 的特征向量构成， V 由 $(P^f)^T P^f$ 的特征向量构成。得最优解为

$$\arg \min \|E(\beta)\| = V \left(\frac{\lambda_i}{\lambda_i^2 + \eta} \right) U^T \Delta d \quad (13)$$

将公式(13)的计算结果代入公式(11)，得到三维数据的重建结果为：

$$S_{new} = \bar{S} + PV \left(\frac{\lambda_i}{\lambda_i^2 + \eta} \right) U^T \Delta d \quad (14)$$

然后为三维重建结果加上纹理图[5]，即完成了人脸的三维重建。

4. 实验结果与分析

4.1. 特征匹配

对目标进行人脸特征匹配时，会出现对无用点的匹配，既浪费时间，也浪费资源，相关数据见表 1。因此在对目标人脸进行人脸识别前，需要图片中的人脸进行截取，为保证还原的保真度，我们采用保持截取人脸的图像尺寸不变的原则。

程序输出得到若干个包含人脸的矩形框位置 (x, y, w, h) 。其中 x 与 y 表示人脸在原图中坐标， w 与 h 表示获取得到的人脸的宽与高。对匹配点的坐标进行适当的线性变换便可得到该特征点在原像素坐标系中的坐标。

截取后，立体匹配区域显然变小，匹配效率自然会大幅度提高。实验表明，从图像中获取人脸区域再进行立体匹配可以极大提高匹配的效率。

特征匹配前使用 SIFT 算法对人脸的特征点进行提取，将特征点的位置画到图像上，提取结果如图 2 所示。

Table 1. Three-dimensional information of five sets of feature points
表 1. 五组特征点的三维信息

数据组	左右匹配点	视差	三维坐标	计算距离/mm	手持激光测距/mm	误差
数据组一	(488, 239) (438, 239)	50	-24.23, -165.96, 2767.60	2772.68	2783.43	0.388%
数据组二	(496, 292) (443, 292)	53	-11.03, -78.17, 2610.95	2612.14	2645.78	1.321%
数据组三	(501, 292) (442, 292)	59	-3.26, -70.22, 2345.43	2346.48	2365.32	0.823%
数据组四	(497, 262) (430, 262)	67	-7.55, -96.94, 2067.38	2067.66	2077.91	0.501%
数据组五	(490, 362) (412, 362)	78	-13.52, 17.23, 1774.11	1774.24	1786.26	0.789%



Figure 2. Results of feature point extraction by SIFT
图 2. SIFT 提取特征

特征点集中分布在脸部边缘、头发边缘、以及眼耳口鼻处。脸部边缘和头发边缘颜色变化大，像素灰度值有突变，所以有很明显的边缘特征。眼耳口鼻处表面形状复杂、凹凸不平，有很多角点、边缘、线段等，在很小的局部区域内表现为颜色、亮度、形状有明显差别。

将两张人脸图片中对应的特征点，使用 BruteForce 进行匹配，绘制从两个图像中匹配出的关键点，匹配结果如图 3 所示。剔除边界特征点的判断方法比较严格，有可能误删正确的特征点，特别是特征点靠近图像外围时。所以，为了减少误删特征点，裁剪脸部图像时，在周围保留了一圈对人脸三维重建影响不大的背景，即墙面。这样的话，误删的匹配点主要是墙面的特征，不影响人脸的特征提取和匹配。

图 3 中特征匹配的结果中，大量的特征点间的线都是水平的。符合实验精度的要求。

4.2. 三维坐标求解

实验是基于 Matlab 和 VisualC++，为了验证本文采用方法的有效性，进行如下实验：随机选取一组图像对，利用公式计算出图中匹配的特征点的三维世界坐标值，如表 1，并通过相关数据的计算比较进行验证。

特征点到摄像机的计算距离与实际距离误差不超过 1.5%。可见，本文三维重建的正确度比较高。所产生的误差由如下几个方面引起：首先，摄像机标定过程中仍然存在不可避免的误差，例如未考虑薄透镜畸变；其次，实际测量的距离信息也可能存在误差；第三，特征点提取存在误差；第四，光线等引起的噪声干扰。

4.3. 人脸三维重建

使用本文中算法,对双目相机获取的人脸图像,进行特征点提取和人脸三维重建。人脸三维重建模型实验结果如图4所示。

选择平均欧氏距离作为客观评价标准 $e_{euc}(S_r, S_t) = \frac{1}{|V|} \sum_{i \in V} \|V_{r,i} - V_{t,i}\|$ 。 S_r, S_t 是两个形状的曲面, V 分别是曲面上的顶点集合。 $V_{r,i}, V_{t,i}$ 分别是 S_r, S_t 上的顶点。三维重建精度与时间如表2所示,选择文献[12]作为对比算法。本文算法的重建精度有所下降,但是其仍可以保持较好的精度,却可以极大地提高三维重建的效率,为实时三维识别奠定了基础。

5. 结语

本文采用 SIFT 算法、特征点标定方法实现了人脸的三维重建。最终的人脸模型网格足够稠密,精度足够高,而且与实际人物相符,准确度也达到了预期的要求。与其它常用的人脸三维重建技术相比,本文提出的方法原理更加简单直观,对设备和环境的要求不高,不依赖标准人脸模型或人脸数据库。建模的准确度、精度和速度都足够满足常规的要求,适用范围很广,在目标识别、动态追踪、三维重建等计算机视觉领域有一定的优势。但 SIFT 算法的计算量很大,现阶段还不能将其直接用于人脸的实时三维重建。进一步提高 SIFT 算法的计算速度是未来的一个研究方向。

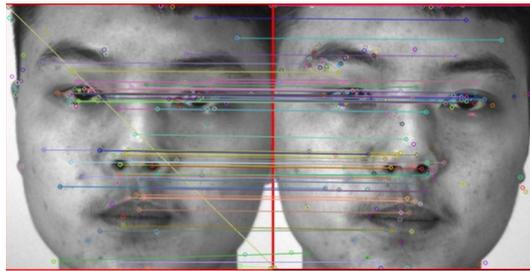


Figure 3. Results of feature matching by SIFT

图3. SIFT 特征匹配



Figure 4. Results of 3D reconstruction of the face

图4. 人脸三维重建结果

Table 2. Contrast results of 3D reconstruction of face and time of use
表2. 重建精度与时间的对比

方法	文献[12]算法	本文算法
欧氏误差/mm	8.34	12.56
重建时间/s	245.23	14.53

基金项目

本文得到国家大学生创新创业训练计划项目(201710490012)、武汉工程大学校长基金项目(2017042)的支持。

参考文献

- [1] Zhang, Q. and Pless, R. (2014) Extrinsic Calibration of a Camera and Laser Range Finder (Improves Camera Calibration). *Proceedings of 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014)*, Sendai, 28 September-2 October 2014.
- [2] Liao, H.B., Chen, Q.H., Zhou, Q.J., *et al.* (2012) Nonuniform Resampling Based on Method for Pixel-Wise Correspondence between 3D Faces. *Journal of Visual Communication and Image Representation*, **23**, 924-931. <https://doi.org/10.1016/j.jvcir.2012.06.005>
- [3] Patel, A. and Smith, W.A.P. (2012) Driving 3D Morphable Models Using Shading Cues. *Pattern Recognition*, **45**, 1993-2004. <https://doi.org/10.1016/j.patcog.2011.11.013>
- [4] Kemelmacher-Shlizerman, I. and Basri, R. (2011) 3D Face Reconstruction from a Single Image. Using a Single Reference Face Shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **3**.
- [5] Patel, A. and Smith, W.A.P. (2012) Driving 3D Morphable Models Using Shading Cues. *Pattern Recognition*, **45**, 1993-2004.
- [6] 谷天, 陈江龙, 李科, 等. 一种用于三维人脸重建的快速稠密视差图生成方法[J]. 西华大学学报(自然科学版), 2016, 35(1): 77-79.
- [7] 隋巧燕, 董洪伟, 刘蕾. 双目下点云的三维人脸重建[J]. 现代电子技术, 2015(4): 102-105.
- [8] 鹿乐, 鞠园, 蔡亮, 等. 基于特征分块的快速三维人脸重建[J]. 火力与指挥控制, 2014(z1): 152-155.
- [9] 孙曦. 基于双目立体视觉的人脸三维重建与识别[D]: [博士学位论文]. 北京: 中国科学技术大学, 2011.
- [10] 王小文. 基于形变模型的多摄像机人脸三维重建[D]: [硕士学位论文]. 北京: 北京理工大学, 2010.
- [11] Xu, G. and Zhang, Z. (2013) *Epipolar Geometry in Stereo, Motion and Object Recognition: A Unified Approach*. Springer Science & Business Media, Berlin.
- [12] 吴子扬, 周大可, 杨欣, 等. 由粗到精的三维人脸稀疏重建方法[J]. 吉林大学学报(信息科学版), 2013, 31(6): 582-588. <https://doi.org/10.3969/j.issn.1671-5896.2013.06.005>

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-6753, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: jisp@hanspub.org