Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Automatic 3D Ear Reconstruction Based on Binocular Stereo Vision

Hui Zeng, Zhi-Chun Mu, Kai Wang, Chao Sun
School of Information Engineering
University of Science and Technology Beijing
Beijing, China
hzeng@163.com

***Abstract*—This paper presents an automatic 3D ear reconstruction method based on binocular stereo vision. At first, we calibrate the stereo vision system by Zhang's method. Then the quasi-dense matching method is performed. We use SIFT feature based matching approach and the coarse to fine strategy to compute the seed matches. The adapted match propagation algorithm with known epipolar geometry constraint is used for obtain quasi-dense correspondence points. Finally the 3D ear model can be reconstructed by triangulation and the results are optimized using the bundle adjustment technique. Extensive experimental results have shown that our proposed method can obtain denser 3D ear model than Liu's multi-view based method. It can get sufficient 3D ear points with lower cost and higher efficiency.**

***Keywords*—3D Ear reconstrction, SIFT, propagation, quasi-dense matching, triangulation, bundle adjustment**

I. INTRODUCTION

Ear is a promising biometric and it has been used as major feature in forensic science for many years. Compared with other popular human features, ear has many advantages, such as it has a rich and stable structure that changes little with age and does not suffer from changes in facial expression and so on [1-3]. Much work has been done in this field and those approaches roughly can be classified into two categories: 2D-based approaches and 3D-based approaches [4-8]. Compared with the 2D-based approaches, the 3D-based approaches are relatively insensitive to pose and lighting variations. So in recent years, more and more researchers began to pay attention to the recognition approaches based on 3D ear data.

For 3D ear recognition, most approaches are based on range data. Pin Yan and Kevin W. Bowyer proposed an automatic 3D ear recognition system, which includes automatic ear extraction, and ICP based 3D shape matching for recognition [5]. Hui Chen and Bir Bhanu proposed a 3D ear recognition method based on local surface patch and ICP method [6,7]. Although these approaches are more robust for pose and lighting variations, they need high-cost facilities and expensive computation, and the data acquisition process requires the user to maintain a relatively still pose for several seconds [9,10]. However, the acquisition of the 2D images is more convenient and the process is nearly real-time and noninvasive. So 3D ear recognition based on 2D images becomes a new research focus in the field of biometrics. How to obtain effective 3D ear model from 2D images is the first key issue in 2D image based 3D ear recognition systems. So in this paper, we discuss the 2D image based 3D ear reconstruction technique.

Heng Liu *et al.* propose a semiautomatic 3D ear reconstruction method using multiple views [8]. It applies multi-view epipolar geometry and motion analysis principles to obtain 3D ear model. At first, they use Harris corner detector and RANSAC technique to detect and match the two-view ear feature points automatically. The results show that this conventional detection and matching method is not suitable for ear images and it only obtain a few right matching feature points in the auricle. Thus they propose a user interactive way based on ear contour division to select and match ear feature points. Their experimental results show that it can obtain only about 300-600 vertices and the points of the reconstructed 3D ear model is sparser than those 3D ear models coming from range scanner and structure light device.

Steven Cadavid *et al.* present two kinds of video based 3D ear reconstruction and recognition approaches using structure from motion (SFM) and shape from shading (SFS) techniques [9,10]. For the approach based on SFM technique, they use the force field transformation method to smooth out the image artifacts firstly. Then the ear region is segmented and the ear curves are detected by the ridges and ravines detection algorithm. The feature points are tracked across the video sequence using the Kanade-Lucas-Tomasi(KLT) Tracker. The factorization method developed by Tomasi and Kanade is used to reconstruct 3D ear model. At last, the postprocessing step is carried out for outlier filtering and the Partial Hausdorff distance (PHD) metric is used to compare the test model and the database ear models. For the approach based on SFS technique, a set of frames is extracted from a video clip and the ear region is segmented firstly. Then the 3D ear models corresponding to each frame are derived using the SFS technique. Finally, the resulting models are aligned using the Iterative Closet Point (ICP) algorithm and cross validation is performed for recognition. Both the above 3D ear reconstruction approaches can obtain the 3D ear models from the video clip automatically, but they are still have some disadvantages. In the SFM approach, it is difficult to obtain sufficient corresponding feature points. The majority of the reconstructed models are so sparse that they can't meet the need of the recognition system. The SFS approach can obtain dense representation of the ear, but it is highly sensitive to

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

lighting variations because it is an ill-posed problem and derives the 3D structure based on illumination and reflectance properties of the scene. The 3D ear representations of the SFS approach are 2.5-D models and the 3D ear information has not been utilized sufficiently.

Because the SFM method is more robust to lighting variations and its accuracy is better than that of the SFS method [8-10], we propose an automatic 3D ear reconstruction approach using the SFM technique. It can obtain quasi-dense 3D ear models and the whole processes are performed automatically. The experimental images are captured by the binocular stereo vision system. Extensive experimental results have testified its efficiency.

The rest of this paper is organized as follows. Section 2 introduces the binocular stereo vision system and the system calibration method. Section 3 describes detail steps of our quasi-dense matching method, including initial seed points extraction and matching, match propagation and 3D ear reconstruction. Section 4 provides the experimental results. Section 5 concludes the paper.

II. SYSTEM CALIBRATION

In this paper, we use the binocular stereo vision system to capture the ear images. As shown in figure 1, the system consists of two Canon EOS 450 digital cameras and a tripod. The key problems that affect the performance of the SFM based 3D reconstruction approaches are how to estimate the camera intrinsic parameters and recover the camera motion accurately and robustly. If the images are taken by a stereo vision system, the “motion” denotes the relative position relationship of the images. In order to improve the accuracy of the 3D ear models, the camera intrinsic parameters and the relative position relationship of the two cameras are fixed in the process of the collection of the ear image database. We use a checker pattern to calibrate the system by Zhang’s method [11].The intrinsic parameters of the two cameras and their relative position can be computed. The epipolar geometry of the two cameras can also be determined [12].

Figure 1. The binocular stereo vision system.

III. QUASI-DENSE MATCHING AND 3D EAR RECONSTRUCTION

First of all, we extract the ear regions from the lateral face images by hand. As shown in figure 2(a) and (b), the green rectangular lines are the margins of the selected ear regions. It is worth noting that the black rectangles are for the purpose of obscurity identity. Here, we can also use other automatic ear region detection algorithms, such as Adaboost algorithm etc.

As the ear images have similar textures, conventional image matching approaches are not suitable for ear images and they can only obtain sparse matching points. In this paper, we use the match propagation algorithm to get quasi-dense correspondences. In order to improve the accuracy and the robustness of the system, we use the SFM technique to reconstruct the 3D ear model. Detail steps are shown as bellows.

A. Seed Matches

In recent years, the local invariant descriptor based matching approaches have made great progress. The researchers have testified that the SIFT (Scale Invariant Feature Transform) has better performance and it is one of the most popular feature for image point correspondences in recent years [13]. It is invariant to image scale and rotation, and distinctive enough for matching. Furthermore, it is robust to affine distortion, change in 3D viewpoint, addition of noise, and change in illumination [14]. So we use the SIFT-based matching algorithm to obtain seed matches in this paper.

At first, the keypoints are obtained by detecting the stable local extreme in DOG (Difference-of-Gaussian) scale space. The locations of the keypoints are refined by performing a detailed fit to the nearby data location, scale, and ratio of principal curvatures. Secondly, the dominant orientation is assigned by forming the orientation histogram from the gradient orientations of the neighborhood of the keypoint. Then compute the gradient magnitude and orientation at each image sample point in a region around the keypoint location and the local keypoint descriptor is constructed. Finally the ratio of closest to second-closest neighbors of each keypoint can be used for matching [13].

As the ear regions have similar texture, it is easy to get false matches. The epipolar geometry constraint based RANSAC (Random Sample Consensus) robust estimation method is the most common used solution for solving this problem [12]. In this paper, as the epipolar geometry has been accurately estimated in the step of system calibration, we only use the epipolar geometry constraint to remove the false matches. If the distance from the keypoint to its corresponding epipolar line is larger than the preset threshold, we believe that this keypoint is the correspondence outlier. After remove all the outliers, the ZNCC (Zero-mean Normalized Correlation Coefficient) based coarse to fine matching strategy is adopted to improve the location accuracy of the matching points. Here we use the method of parabolic interpolation to refine the point to sub-pixel accuracy. After the above steps, we can get seed matches accurately and robustly. The results are shown in Figure 2 (c) and (d), and the final seed matches are marked with green points.

B. Propagation Algorithm

The feature based matching algorithms, such as Harris and SIFT, only obtain sparse matches. In the application of the 3D reconstruction system, sparse correspondence can describe the rough contour and the details can’t be presented. The dense matching algorithms can obtain dense correspondences, but they are expensive in terms of time and memory. The quasi-dense approached are developed to over come these deficiencies. They are sufficient for 3D reconstruction and more efficient than the dense approaches [15,16]. There are two kinds of approaches, including the region based propagation approach and the point based propagation approach. In this paper, we use the point based propagation approach. It is adapted from the approach proposed in [17]. The global best first strategy is adopted in propagation. It relies on only a few most reliable seed points, has good stability in low-textured scenes and can handle half-occluded areas. Because the cameras have been calibrated beforehand, the epipolar geometry is known. So we can use the epipolar geometry constraint to reduce the research to one-dimension instead of two.

At each step of the propagation, the best seed matches are selected using ZNCC measure and SIFT descriptor firstly. Then we look for new potential matches using 1D disparity gradient limit constraint along the epipolar lines, intensity similarity constraint, confidence constraint and uniqueness constraint. Finally the best matches are added to the current list of seed matches. In order to make the correspondences distribute evenly and improve the location accuracy to sub-pixel, the resampling step is performed. Figure (e) and (f) are two quasi-dense matching results. Here the quasi-dense correspondences are marked with green points. We can see that the proposed propagation algorithm can obtain sufficient matching points effectively.

In this section, we describe our adapted match propagation algorithm with known epipolar geometry constraint. Compared with the propagation algorithm described in [17], our propagation algorithm use both ZNCC measure and SIFT descriptor to compare the similarity of the candidate seed matches. It is more robust and has better stability, especially for ear images that have a great deal of similar textures. In the process of propagation, the known epipolar geometry constraint can improve its performance in time and memory.

C. 3D Ear Reconstruction

After the quasi-dense matching points have been extracted, we can estimate their corresponding 3D points using triangulation principle. It belongs to the SFM method. As the system has been calibrated, the camera matrix P_1 and P_2 are known. For each point correspondence $m_1 \leftrightarrow m_2$, the 3D points X_i can be computed from the following equations

$$s_1 m_1 = P_1 X_i \quad s_2 m_2 = P_2 X_i$$

Where s_1 and s_2 are the non-zeros scale factors.

In real applications, the system calibration results may have errors and the correspondences are noisy inevitably, so we must optimize the 3D reconstruction results using bundle adjustment technique. Here we use L-M algorithm to minimize the Sampson geometric function. After the 3D reconstruction points have been computed, they can be meshed and visualized using view-dependent texture mapping technique. Figure 2 (g) and (h) show the 3D reconstruction points under different viewpoints, (i) is the mesh result and (j) texture mapping result. From these figures we can see that the proposed quasi-dense 3D ear reconstruction can be computed effectively.

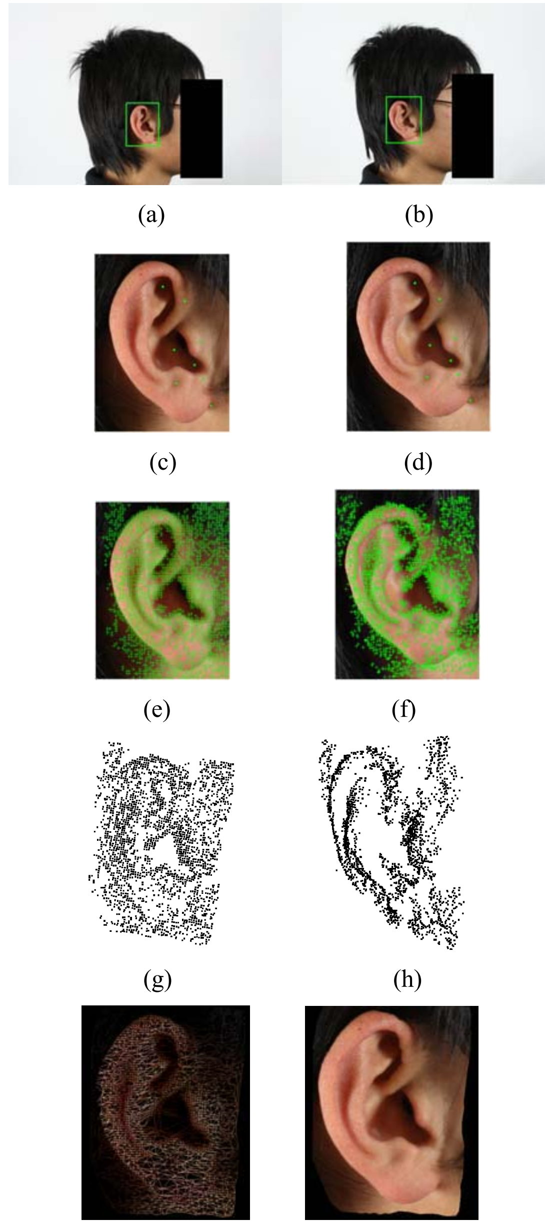

Figure 2. 3D ear reconstrction from binocular stereo images, (a) and (b) are two lateral face images, (c) and (d) are the seed matching results of the two corenspoing ear imges, (e) and (f) are their quais-dense matching results, (g) and (h) give the 3D reconstructed points under different viewpoints, (i) is the mesh of the 3D reconstructed points and (j) showns the texture mapped 3D result.

IV. EXPERIMENTAL RESULTS

Our experiments are carried out on the USTB ear database. It consists of 1096 images, 8 images for each of 137 subjects. For each subject, we capture 2 pairs of left lateral face images and 2 pairs of right lateral face images. The images are captured two Canon EOS 450 digital cameras. The resolution of the image is 4272×2848 and the size of the ear region is about 500×700. The distance from the ear to the camera is about 1m. At first, we calibrate the system using a 50mm× 50mm checker pattern. The camera matrixes and the epipolar geometry of the binocular stereo vision system can be computed. The mean reprojection error of the pattern corners is 0.15 pixels. Then quasi-dense matching algorithm is performed to obtain sufficient correspondence points. Finally the 3D ear points can be reconstructed using triangulation principle and the results are optimized by bundle adjustment. The Figure 2 shows detail steps of 3D ear reconstruction. The 3D ear points are computed from two right lateral face images. Figure 3 gives another 3D ear reconstruction example computed from two left lateral face images.

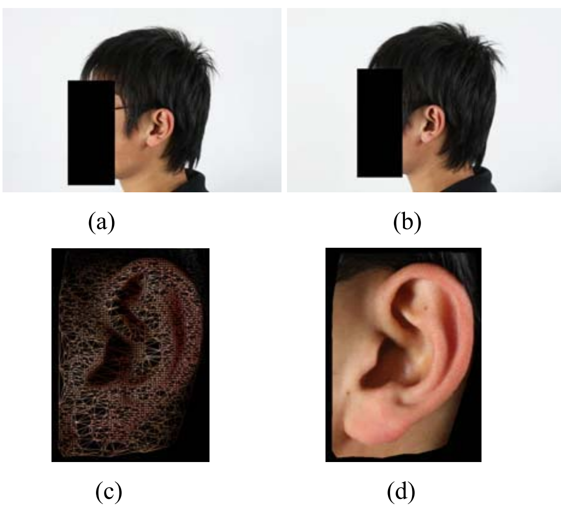

Figure 3. An example of 3D ear reconstrction result from binocular stereo images, (a) and (b) are two left ateral face images, (c) is the mesh of the 3D reconstructed points and d) showns the texture mapped 3D result.

For each pair stereo images, we can obtain 2000-3000 points. TABLE 1 gives the comparison of different methods. Although the number of our 3D reconstruction vertices is smaller than that obtained by range scanner, it is almost the same as structure light and denser than the semi-automatic method proposed in [8]. So our quasi-dense reconstruction method can obtain relative denser 3D ear points with low cost.

TABLE I. COMPARISION OF DIFFERENT METHODS

	Methods			
	Range Scan	***Structure Light***	***Multi-view***	***Quasi-Dense***
Vertices	7000-9000	2000-4000	300-600	2000-3000

V. CONCLUSION

In this paper, we present an automatic 3D ear reconstruction method based on binocular stereo vision. For ear images, feature based matching method can only obtain sparse correspondence points and dense matching methods are expensive in terms of time and memory, so we adopt the quasi-dense matching method. The 3D points are computed by the SFM method that has higher accuracy and better robustness than the SFS method. Based on the 3D ear models computed by our quasi-dense method, we will build an automatic 3D ear recognition system.

ACKNOWLEDGMENT

This paper is supported by: (1) the National Natural Science Foundation of China under the Grant No. 60573058; (2) the key discipline development program of Beijing Municipal Commission of Education under the Grant No. XK100080537.

REFERENCES

[1] D. J. Hurley, B. Zrbab-Zavar and M. S. Nixon, The Ear as a Biometric, Handbook Of Biometrics, pp. 131-150

[2] M. Burge and W. Burger. Ear biometrics. In A. Jain, R. Bolle, and S. Pankanti, editors, BIOMETRICS: Personal Identification in a Networked Society, pp:273-286. Kluwer Academic, 1998.

[3] D. Hurley, M. Nixon, and J. Carter. Force Field Energy Functionals for Ear Biometrics. Computer Vision and Image Understanding, vol.98, pp.491-512,2005.

[4] Christopher Middendorff, Kevin W. Bowyer and Ping Yan, Multi-Modal Biometrics Involving the Human Ear, IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-2, June 2007.

[5] Ping Yan and Kevin W. Bowyer. Biometric Recognition Using 3D Ear Shape, IEEE Transactions on Pattern Analysis And Machine Intelligence, vol. 29, no. 8, pp. 1297-1308, August 2007.

[6] Hui Chen and Bir Bhanu, Human Ear Recognition in 3D. IEEE Transactions on Pattern Analysis And Machine Intelligence, vol. 29, no. 4, pp. 718-737, April 2007.

[7] Hui Chen and Bir Bhanu. Efficient Recognition of Highly Similar 3D Objects in Range Images, IEEE Transactions on Pattern Analysis And Machine Intelligence, vol. 31, no. 1, pp. 172-179, January 2009.

[8] Heng Liu, Jingqi Yan and David Zhang, 3D Ear Reconstruction Attempts: Using Multi-view, International Conference on Intelligent Computing, pp. 578-583, 2006.

[9] Steven Cadavid and Mohamed Abdel-Mottaleb, 3-D Ear Modeling and Recognition From Video Sequences Using Shape From Shading, IEEE Transactions on Infoemation Forensics And Security, vol. 3, no. 4, pp. 709-718, December 2008.

[10] Steven Cadavid and Mohamed Abdel-Mottaleb, Human Identification Based on 3D Ear Models, First IEEE Inrenational Conference on Biometrics:Theory, Applications, and Systems, pp.1-6, Sept. 2007.

[11] Zhengyou Zhang, Flexible Camera Calibration by Viewing a Plane from Unknown Orientations, The proceedings of the seventh IEEE Inrenational Conference on Computer Vision, pp. 666-673, vol. 1,1999.

[12] R Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press 2000.

[13] D. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, 2004, 60(2), pp.91–110.

[14] K. Mikolajczyk and C. Schmid. A Performance Evaluation of Local Descriptors, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 27(10), pp.1615–1630.

[15] Lhuillier Mand Quan L. Robust dense matching using local and global geometric constraints, In Proceedings of the 16th International Conference on Pattern Recognition, Barcelona, Spain, 2000, vol(1), pp.968–972.

[16] Lhuillier M and Quan L. Match propagation for image-based modeling and rendering, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(8):1140–1146.

[17] LhuillierMand Quan L. A quasi-dense approach to surface reconstruction from uncalibrated images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3):418–433