반 밀집 정합에 기반한 저가형 3차원 주름 데이터 복원☆

A Low Cost 3D Skin Wrinkle Reconstruction System Based on Stereo Semi-Dense Matching

장 천*　　　황보 택 근**
Qian Zhang　　Taeg-Keun Whangbo

요　약

본 논문은 스테레오 영상에 기반한 3차원 주름 데이터 복원 시스템을 제안한다. 일반적으로, 3차원 데이터 복원 연구는 스테레오 영상 또는 비디오 영상을 통해, 문화유산과 같은 건물이나 주변 환경에 적용 하는 연구가 최근 활발히 진행되고 있다. 3차원 데이터 복원에서 사물 측정의 목적은 각 영상 간의 깊이 정보 계산을 통해 3차원 데이터를 획득하는 것이다. 본 연구를 진행하기 위해선 몇 가지 고려해야할 사항이 있다. 첫째로, 카메라 성능과 비 균일한 구성의 피부, 그리고 조명의 영향 때문에 촬영 시 주름의 완전한 정보를 얻기가 힘들다. 따라서 본 논문은 주름의 정보를 최대한 완전하게 얻기 위해, 근접 촬영이 가능한 접사렌즈를 사용하였다. 둘째로, 피부 영상은 영상 분할이나 특징점을 추출하는데 부정확한 문제점이 있어, 스테레오 영상의 밀집 정보를 얻기가 힘들다. 따라서 본 논문은 주름의 정확한 깊이 정보 계산을 위해 반 밀집에 기반한 스테레오 정합 알고리즘을 사용한다. 본 시스템은 기존의 3차원 스캐너와 비교해 비용 절감의 효과가 있으며, 실험 결과 일반적인 모델링 방법보다 고성능의 결과를 보여주며 이를 통해 활용방안이 많을 것으로 예상된다.

Abstract

In the paper, we proposed a new system to retrieve 3D wrinkle data based on stereo images. Usually, 3D reconstruction based on stereo images or video is very popular and it is the research focus, which has been applied for culture heritage, building and other scene. The target is object measurement, the scene depth calculation and 3D data obtained. There are several challenges in our research. First, it is hard to take the full information wrinkle images by cameras because of light influence, skin with non-rigid object and camera performance. We design a particular computer vision system to take winkle images with a long length camera lens. Second, it is difficult to get the dense stereo data because of the hard skin texture image segmentation and corner detection. We focus on semi-dense stereo matching algorithm for the wrinkle depth. Compared with the 3D scanner, our system is much cheaper and compared with the physical modeling based method, our system is more flexible with high performance.

☞ KeyWords: 3D reconstruction, Skin Wrinkle, Semi-dense matching, Image rectification, Light compensation,
3차원 복원, 피부 주름, 반 밀집 정합, 이미지 보정, 빛 보정

1. Introduction

Nowadays there is more and more demand for 3D models in computer graphics, virtual reality and communication. For the application in medical analysis, both image-based and modeling-based skin analysis, the 3D skin model is very popular and flexible involving computer assisted diagnosis for dermatology, topical drug efficacy testing for the pharmaceutical industry, and quantitative product comparison for cosmetics. For the computer animation, skin wrinkle will bring abundant expression to the actor. Quantitative features of skin

* 정 회 원 : 경원대학교 일반대학원 전자계산학과 박사과정
aazhqg@hotmail.com
** 종신회원 : 경원대학교 IT대학 부교수
tkwhangbo@kyungwon.ac.kr(교신저자)
[2008/10/30 투고 - 2008/11/3 심사(2009/02/26 2차) - 2009/03/19 심사완료]
☆ 본 연구는 중소기업청에서 시행한 산학협력실 지원사업 (과제번호 : 2007-0152)의 연구결과로 수행되었음

surface are the significant but difficult task. In current methods, information has been limited to the visual data such as images, video, etc. The 3D features of skin can make the computer-based visual inspection more accurate. Our work focuses on skin wrinkle 3D data retrieval from image sequences.

There are techniques with satisfying results for microscopic objects, others for small, medium and large objects. These techniques include laser scanning techniques, shape from stereo, and shape from video [1] or shading [2], and so on. Laser scanning techniques are based on a system with a laser source and an optical detector. The laser source emits light in the form of a line or a pattern on the objects surface and the optical detector detects this line or pattern on the objects. Through well-know triangulation algorithm the system is able to extract the geometry of the objects. The most important advantage of laser scanners is high accuracy in geometry measurement. Also it has some disadvantages. First geometry can be extracted without any texture information. Second the high cost of hardware components includes laser, the light sensor and the optical system. Third it is practical impossible to stay immobile for some seconds scanning, such as breath and wink. The technique shape from stereo is the extrapolation of as much geometry information as possible from only a pair of photographs taken from known angles and relative positions, which is the simulation of human visual system. Calibration is critical in terms of achieving accurate measurements. The method can either be fully automated or manually operated. Advantages of this method are the ability to capture both geometry and texture, the low cost and portability. A disadvantage of the method is its low resolution [3].

The skin surface is a complex landscape influenced by view direction and illumination direction [4].That means we can take skin surface as a type of texture, but this texture is strongly affected by the light and view direction, and even the same skin surface looks totally different. Lots of Previous work has been done for extracting 3D texture and shape from image sequences. Among them, shape from shading [5] and photometric stereo [6] are effective approaches for a surface reconstruction. Both techniques use the pixelwise in the input images. They need an integration step to reconstruction a surface. This step tends to accumulate errors from pixel to pixel. That means the final relative height differences between distant points may not come out very accurately. Another type of approach is based on multi-view stereo matching. As usual, it is applied for culture heritage, building and other scene reconstruction [7] [8] [9]. The user acquires the images by moving one camera or two cameras together around an object or scene. Neither the camera motion nor the camera settings have to be known a priori. In our work, we try to apply this method to recover the skin wrinkle. There are several challenges in our research. First, it is hard to taking the full information wrinkle photos by cameras because of light influence, skin with non-rigid object and camera performance. Second, it is difficult to get the dense stereo data because of the hard skin texture image segmentation and corner detection. In the paper, a particular computer vision system was proposed to acquire the image pair. For the sparse stereo matching, we detected the corresponding points manually, and refined them through fundamental matrix from camera calibration. Considering the difficulty in dense stereo matching, we applied the semi-dense stereo matching algorithm to acquire more accurate corresponding points. The approaches presented here combines many ideas and algorithm. We bring them together in order to retrieve skin wrinkle which is the detail from object surface accurately.

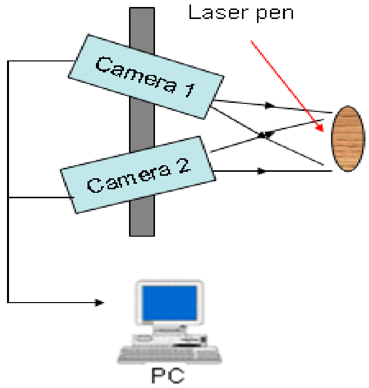
This paper was organized as follows: section 2 introduces the system construction, which includes system design, some preprocessing parts such as light compensation, image rectification, etc. We describe the detail of semi-dense stereo matching algorithm in section 3, and this is the significant step in 3D skin wrinkle retrieval. Experiment result is shown in section 4. We summarize our work in section 5, also includes the future works.

2. System Construction

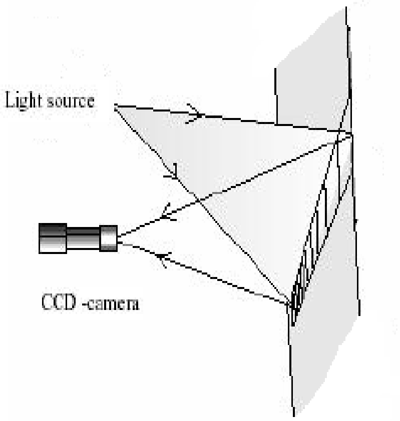
In our system, we take wrinkle images with two cameras at the same time because of the non-rigid property of wrinkle. In order to get dense matching data, we utilize the affine geometry, projective geometry and epipolar geometry after we finish the precise camera calibration and stereo calibration. Also we should pay attention to the wrinkle image taken experiment and image preprocessing.

2.1 Images Acquisition System Design

Our test images taken distance will be controlled in 18~25cm in order to get clear wrinkle photos with skin texture details. In the close distance, long focus lens will be necessary. For controlling two cameras to take photos at the same time, computer is needed with camera SDK improvement. It is hard to make two cameras focus on the same position on the skin with wrinkle. We design a special equipment to set the cameras at the same height, which shows in figure 1. It helps us to obtain wrinkle images more conveniently and stably. A laser pen is applied helping camera to find the same part in the skin. We choose table lamp with parallel light in the dark room without camera light. Our computer vision system structure for taking images is shown in the Figure 1 and Figure 2 respectively.



(Figure 1) Computer Vision System Structure



(Figure 2) Images Acquisition System

2.1 Image Preprocessing: Light Compensation and Image Rectification

In our system, we use chess template for camera calibration. We do the single camera calibration one by one to calculate the camera intrinsic and camera extrinsic matrix. The camera intrinsic matrix includes camera focus and projection center coordinate. The camera extrinsic matrix includes rotation matrix and transformation vector. Before we do the stereo camera calibration, affine transformation will be used for template rectification. Then we can calculate the two cameras' position matrix: rotation matrix and transformation matrix. All of the parameter matrix will be applied for computing Fundamental matrix which is the important and basic correlation between the two image corresponding points.

2.1.1 Light Compensation

Generally skin images can be totally different even by a slight different position because there are influenced highly by light. In our images acquisition system, the luminance of skin wrinkle images is very low. We couldn't recognize skin wrinkles from the dark images directly. Light compensation algorithm is necessary. We transform each pixel from RGB space to YCbCr space, normalize each pixels gray level to a new value through approximating a second-order equation to mark all pixels' gray scale, and make a range of the transformed gray level between 0 and 255 with a scale factor. The detail of the light compensation approach is referred to paper [10].

2.1.2 Image Rectification

Feature detection for the skin wrinkle images is important before image rectification. Corner detection is the usual algorithm for feature detection in the image. But in our images, the skin texture with several light influences is not the real detail structure. Corner detection and other feature detection algorithm, such as SIFT (Scale-invariant feature transform) couldn't work. In our system, we apply the color model to separate hair and skin on the wrinkle images. Because they have different color: black and skin color. At least 8 corresponding points should be selected for image rectification manually from the hair left images.

Rectification is a process used to facilitate the analysis of a stereo pair of images by making it simple to enforce the two view geometric constraint. For a pair of related views the epipolar geometry provides a complete description of relative camera geometry. Once the epipolar geometry has been determined it is possible to constrain the match for a point in one image to lie on the epipolar line in the other image. The process of rectification makes it very simple to impose this constraint by making all matching epipolar lines coincident and parallel with an image axis.

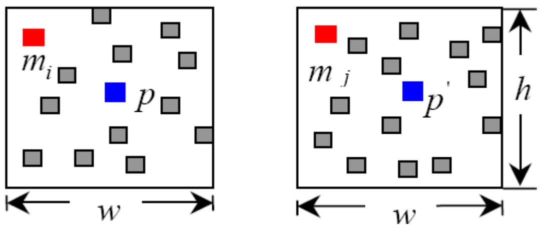
Actually planar rectification is not general and if there is a large forward component in the camera movement it may produce unbounded, large or badly warped images. It was not a problem because stereo vision was usually performed using stereo rigs with near parallel cameras. In our system, the eye wrinkle real size is about 3*3cm and the object distance will be controlled in 18~25cm. Two cameras couldn't focus on the same skin region while they paralleled each other. In our system, first existing matches between the images will be used to calculate the epipolar geometry in order to determine a rectification which minimizes perspective distortion effects such as motion parallax between the images. Secondly, we use the same nonlinear transformation [11] for both images, and it overcomes a problem with existing general techniques where the application of different nonlinear transformations to each image results in matching image features being warped differently in each image. Finally, we simplify the approach of [11] and handle a number of other minor problems overlooked by previous work such as sub pixel coordinates and infinite epipoles.

3. Semi-Dense Matching

For object 3D reconstruction and modeling accurately, the best way is the dense stereo matching algorithm. Dense matching is a theoretically perfect scheme; in practice it is usually inevitable to obtain false matches which are fatal to 3D reconstruction. In the skin wrinkle reconstruction, as we known, it is really difficult to do dense stereo matching though area-based dense stereoscopic matching algorithm. It is a very arduous task to separate the foreground from background in the skin surface. Semi-dense matching [12] [13] was considered to complete the stereo matching in our project. Actually, it is a tradeoff between the sparse matching and the dense matching. It works on more point pairs than sparse matching, and more accurate point pairs than dense matching.

3.1 Seeds and the propagating areas

Sparse matching was done before image rectification and also the matching points were refined after image ratification. In the neighborhoods of a pair of matched points, there should exit some other pairs of matching points that are to some degree related to the matched pair. In the paper, these initial matching points were called seeds. They are representatives of local features. And the process to get new pairs of matching points is called propagation. Figure 3 illustrate the seed and propagating area in the paper.



(Figure 3) The seed and the propagating areas. Image A was from left camera and Image B was from right camera. The p and its corresponding point p' are the seed. Point mi and mj are the candidate point in the propagating area with size w*h.

Each seed has two features: the disparity and reliability. In figure 3, point p is in image A with the coordinates (u, v) and it's matched point p' is in image B with coordinates (u', v'). The disparity ρ is the difference of coordinates between the two control points, $\rho = p' - p = \begin{bmatrix} u' - u \\ v' - v \end{bmatrix}$. A pair of matched points (p, p') should satisfy the epipolar constraint and the correlation constraint within a fixed tolerance. The reliability φdefined below is used to measure quantitatively the fitness of matching.

$$\varphi(r, d, d') = r \times f(\sqrt{d^2 + d'^2})$$

$$f(x) = \begin{cases} 1 - \dfrac{x}{\sigma} & \text{if } 0 \le x \le \sigma. \\ 0 & otherwise \end{cases} \qquad (1)$$

Where, r is the zero-mean normalized cross-correlation measuring for n×n windows around the points p and p' respectively, is a threshold to describe the extent satisfying the epipolar constraint. d and d' are the distance of points p and p' to the corresponding epipolar lines.

In each propagating area, the expectant number of corner points is set in a proportion to its area. With this strategy, only prominent corners in texture-rich areas are extracted, but in areas lacking of texture, non-prominent corners are also extracted. These evenly distributing corners make it possible to obtain matching points distributed evenly across image.

3.2 Matching

For skin image, objects are pores and wrinkles which cross each other and form the basic skin structure called grid texture. The feature between two different view images is the basic skin grid. In the matching part, our algorithm for matching point is based on the corner detection in the related propagating area. Harris detector was used in the corner detection step. Suppose the number of initial seeds is n, than we can find more than 2*n matching points in the propagating area. That means we have at least 2*n seeds now. We continue looking for lots of seeds in next iteration. The stop condition relies on a threshold of similarity between the matching points. The similarity is another important criterion in matching process. Traditionally, correlation is carried out on windows centered at two points respectively. Here, we refer to scheme in paper [12] to calculate the similarity as follows:

Step 1 : Choose $w \times h$ area around each seed as propagating area.

Step 2 : Do corner detection in the propagating area. Along the corresponding epipolar line, look for the candidate matching points.

Step 3 : Calculate the similarity between the candidate matching points. Select the matched points.

Step 4 : New group of seeds acquired to repeat step 1~3, until the similarity is larger than the threshold.

We optimize our algorithm according to the paper [12]. In every iterative process, the newly matched points from the more reliable control points will inhibit the propagation of less reliable ones. We will choose the pair of points with the highest correlation value first to avoid the ambiguity.

4. Experiment

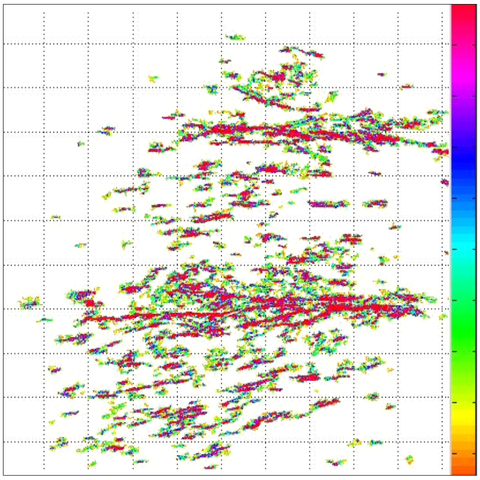
In our system, we use two Cannon 30D cameras with Cannon long focus lens 90mm to take eye folds images. The original image size is 1104*1026. Our system performs on the personal computer with Intel Core 2Duo 2.66GHz CPU and 2G memory. The images with light compensation were shown in figure 4 in the following. After image rectification, our result was shown in figure 5.

(Figure 4) Images after light compensation. Image A is from left camera and image B is from right camera.

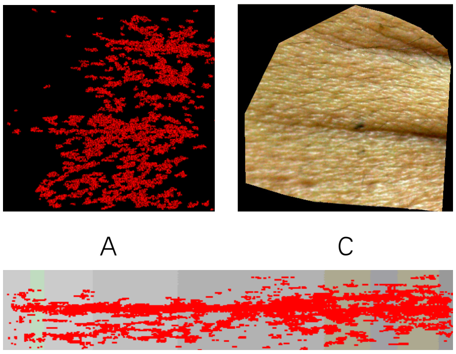
(Figure 5) Images after image rectification. Image A is from left camera and image B is from right camera.

In semi-dense matching part, we set the parameter for corner detection is 0.005. We show the disparity image in figure 6. The number of selected points is 40,578, which is the 3.58% from the total candidate points 1104*1026. The same focus area in the two original images is only one part. The 3D wrinkle data and its model after texture mapping are shown in figure 7. We show another group results in figure 8. There are 11.3475% data from the total pixels in the image matched in this experiment.

In order to choose a suitable technique for 3d object digitization, it should be noted that techniques with satisfying results only for one kind of objects. Paper [3] summarized 'Nine-Criteria' as the possible parameters for choosing a 3D digitization system.

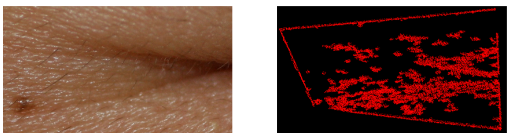


(Figure 6) The disparity image of skin wrinkle. The color from deep red to orange means the depth from top to bottom.



B

(Figure 7) The experiment result. Image A is the 3D eye wrinkle data, image B is point clouds model displayed from a side, and image C is the wrinkle model after texture mapping.



A B

(Figure 8) The experiment result. Image A is from the left camera and image B shows the 3D data model.

In our case, criteria such as cost, portability of equipment, texture acquisition, accuracy of the system, and skill requirements should be considered. For eye wrinkle reconstruction, 3d scanning process will be influenced by eye winking, and 3d scanner could get geometry and texture information simultaneously. Our systems main advantage is the low cost. For example, the price of a 3d portable scanner especial for micro surface of objects is about 30,000 U.S. dollars, and the cost of our system is about 3,000 U.S. dollars includes two cameras with long focus lens, calibration template, and the facility. The disadvantage is our system with low accuracy compared with 3d laser scanner.

5. Conclusion and Discuss

In order to obtain the precise 3D eye wrinkle data without 3D scanner, we proposed a 3D reconstruction system based on images sequences. The human skin has the own specification properties, such as non-rigid object, tiny skin texture structure. We design a computer vision system to take eye wrinkle images by two ordinary cameras with long focus lens. The algorithm for 3D reconstruction comes from the computer vision camera model and projective model. Semi-dense matching algorithm works for wrinkle depth. Human skin image has closer color and blurred skin texture with light influence, our algorithm relies on the more accuracy camera calibration result and stereo calibration result. We need to finish anautonomy process for them with a chess template.

There are still problems in our experiment. We couldn't obtain the ground truth data without 3D scanner. The accuracy of our system couldn't be quantified and compared with other algorithm. In another hand, our approach for 3D wrinkle data retrieval can be applied in computer animation for actor skin expression. In the current, the number of data in point clouds model is not enough to complete mesh model actually. In the future, we will continue focus on the system evaluation and 3D wrinkle modeling.

Reference

[1] A. Chiuso, H. Jin, P. Favaro, and S. Soatto, 'MFm: 3-D Motion and Structure from 2-D Motion Causally Integrated Over Time Implementation', Proceedings of ECCV2000, pp.734-750, 2000.

[2] R. Zhang, P. S. Tsai, J. E. Cryer, and M. Shah, 'Shape from shading: a survey', IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, No. 8, pp.690-706, 1999.

[3] G. Pavlidis, A. Koutsoudis, F. Arnaoutoglou, V. Tsioukas, and C. Chamzas, 'Methods for 3D digitization of Cultural Heritage', Journal of cultural heritage No. 8, pp.93-98, 2007.

[4] O. G. Cula, K. J. Dana, F. P. Murphy, and B. K. Rao, 'Skin Texture Modeling', International Journal of Computer Vision, Vol. 62, No. 1-2. pp.97-119, 2005.

[5] B. K. P. Horn, M. J. Brooks, 'The Variational Approach to Shape from Shading', Computer Version, Graphics & Image Processing, Vol. 33, No. 2, pp.174-208, 1986.

[6] Y. Z. Yu, J. T. Chang, 'Shadow Graphs and 3D Texture Reconstruction', International Journal of Computer Vision, Vol. 62, No. 1-2, pp.35-60, 2005.

[7] M. Pollefeys, L. V. Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch, 'Visual Modeling with a Hand-Held Camera', International Journal of Computer Vision, Vol. 59, No. 3, pp.207-232, 2004.

[8] G. Michael, S. Noah, C. Brian, H. Hugues, and M. S. Steven, 'Multi-View Stereo For Community Photo Collections', Proceedings of ICCV2007, pp.1-8, 2007.

[9] M. Pollefeys, D. Nister, J. M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S. J. Kim, P. Merrell, C. Salmi, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewenius, R. Yang, G. Welch, and H. Towles, 'Detailed Real-Time Urban 3D Reconstruction From Video', International Journal of Computer Vision, Vol. 78, No. 2, pp.143-167, 2008.

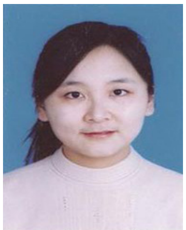
[10] S. I. Choi, C. H. Kim, and C. H. Choi, 'Shadow Compensation in 2D Images for Face Recognition', Pattern Recognition, Vol. 40 , No. 7, pp.2118-2125, 2007.

[11] M. Pollefeys, R. Koch, and L. V. Gool, 'A Simple and Efficient Rectification Method for General Motion', Proceedings of ICCV99, pp.496－501, 1999.

[12] O. Veksler, 'Dense Features for Semi-Dense Stereo Correspondence', International Journal of Computer Vision, Vol. 47, No.1-3, pp.247-260, 2002.

[13] J. Cech, and R. Sara, 'Efficient Sampling of Disparity Space for Fast and Accurate Matching', IEEE Conference on Computer Vision and Pattern Recognition. pp.1-8, 2007.

◐ 저 자 소 개 ◑



장 천(Qian Zhang)
2003년 산동대학교 산업공학과 졸업(학사)
2006년 산동대학교 일반대학원 전자계산학과 졸업(석사)
2006년 ~ 현재 경원대학교 일반대학원 전자계산학과 박사과정
관심분야 : 영상처리, 패턴인식, 얼굴인식
E-mail : aazhqg@hotmail.com

황보 택 근(Taeg-Keun Whangbo)
1983년 고려대학교 공과대학 졸업(학사)
1987년 CUNY 전산학과 졸업(석사)
1995년 Stevens Institute of Technology 전산학과 졸업(박사)
1997년 삼성종합기술원 선임연구원
1997년 ~ 현재 경원대학교 IT대학 부교수
관심분야 : 영상처리, 패턴인식, 컴퓨터그래픽스, 3D 게임엔진
E-mail : tkwhangbo@kyungwon.ac.kr