Two Stages Stereo Dense Matching Algorithm for 3D Skin Micro-surface Reconstruction

Qian Zhang and TaegKeun Whangbo

Department of Computer Science, Kyungwon University,
Sujung-Gu, Songnam, Kyunggi-Do, Korea
aazhqg@hotmail.com,
tkwhangbo@kyungwon.ac.kr

Abstract. As usual, laser scanning and structured light projection represent the optical measurement technologies mostly employed for 3D digitizing of the human body surface. The disadvantage is higher costs of producing hardware components with more precision. This paper presents a solution to the problem of in vivo human skin micro-surface reconstruction based on stereo matching. Skin images are taken by camera with 90mm lens. Micro skin images show texture-full wrinkle and vein for feature detection, while they are lack of color and texture contrast for dense matching. To obtain accurate disparity map of skin image, the two stages stereo matching algorithm is proposed, which combines feature-based and region-based matching algorithm together. First stage a triangular mesh structure is defined as prior knowledge through feature-based sparse matching. Region-based dense matching is done in corresponding triangle pairs in second stage. We demonstrate our algorithm with active skin image data and evaluate the performance with pixel error of test images.

Keywords: Stereo vision, Dense matching, Feature-based, Region-based, Skin reconstruction.

1 Introduction

In medical analysis both image-based and modeling-based skin analysis, 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. Quantitative features of skin 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 and vivid. Our research focuses on human skin micro-surface reconstruction.

As we know, skin surface is a complex landscape influenced by view direction and illumination direction [1]. 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. Some special hardware and techniques (e.g., photometric stereo, reflectance measurement setup) are required to support research. To reduce the impact of these characteristics and also to improve skin data accuracy,

S. Boll et al. (Eds.): MMM 2010, LNCS 5916, pp. 25–34, 2010.

micro skin images are considered in our 3D skin reconstruction. When we observe the skin surface from microcosmic point of view, light projected onto skin surface becomes on part of the skin color with tiny difference from narrow baseline stereo. We will analyze these influences in the paper.

Stereo dense matching between skin image pairs is another difficulty, when we recover skin shape through binocular stereo vision. Corner detection is feasible for feature selection in micro skin images with texture-full wrinkle and vein. It is possible to extract the basic skin structure called as grid texture which is generated by wrinkle and pores. In sparse matching, RANSAC algorithm [2] is a good tool to refine matched points. If we take these sparse matched points as correspondence seeds and find a semi-dense or dense map through seed-growing algorithm, there are many wrong matching because of absence of color and texture contrast in skin images. Even we set the tight constraints to seed-growing and only 2~3 times the number of seed data were matched. We couldn't find the dense map. In region based matching algorithm, energy function was designed and comparison window size was chosen to complete a dense matching through energy function minimization. Graph-cut optimization [3], dynamic programming [4], and region growing [5] are the famous algorithms in this filed. But these methods all rely on a disparity-based formulation. The max disparity will be known as a prerequisite. In our research, we combined the feature based and region based algorithm together to design a two stages matching algorithm to find a dense map. First stage a triangular mesh structure is defined as prior knowledge through feature-based sparse matching. Region-based dense matching is done in corresponding triangle pairs in second stage. It works better than only relying on feature based algorithm and only region based algorithm.

The remainder of this paper is organized as follows. After reviewing related work in Section 2, our matching algorithm is detailed in Section 3, and also the theoretical basis is explained in this section. Experiment results with vivo skin images and the evaluation are shown in Section 4. We conclude our research in Section 5.

2 Related Works

There are several 3D data acquisition techniques with satisfied performance for microscopic objects, small, medium and large objects. These techniques include laser scanning techniques, shape from stereo, and shape from video [6] or shading [7], 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-known 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 practically 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 [8]. In our case, the image resolution of region of interest on the skin is 516*468, while on the same region there are only about 800 data obtained by VIVID 9i laser scanner. Even with high precision sparse stereo matching and camera calibration, we can deduce the difference significantly. In our research, we want to obtain more 3d data by dense matching.

3D reconstruction method based on multi-view stereo matching is widely applied for culture heritage, building and other scene reconstruction [9] [10] [11]. The user acquires the images by moving one camera or two cameras together around an object or scene. Many researchers have been interested in this problem and have proposed different processing pipeline, such as space carving [12], level sets [13], and a discrete labeling formulation through global optimization using graph-cut, belief propagation, dynamic programming, and so on. Fan [14] proposed a mesh-based coding method for surface modeling. They obtain a discrete disparity map using block-based hierarchical disparity estimation and then model the obtained disparity map using the Delaunay triangulation on a set of nodes. If the disparity map is not smooth enough, or if some disparity discontinuities exist, a large number of triangulation nodes must be placed very densely around the noise area to reduce the error. Quan [15] proposed a 3D surface reconstruction algorithm both with and without the epipolar constraint in which images play a symmetric role. The idea is based on seed grow contiguous component in disparity space. The greediness algorithm may cause a complete failure in the repetitive texture in the scene, such as our case, the skin texture images. Later, Jac ech [16] proposed a fast and accurate matching algorithm to find a semi-dense disparity map. Our motivation is from their researches to yield a dense matching. We construct sparse matching nodes for triangulation, and region based global optimization is applied for dense matching within each segmented region.

3 Matching Algorithm

3.1 Overview of Proposed Algorithm

We outline our stereo matching processing in figure 1 shown as follows. Our matching algorithm is suit for rectified image pairs. The rectification algorithm is referred to paper [17]. There are two stages included in our pipeline:

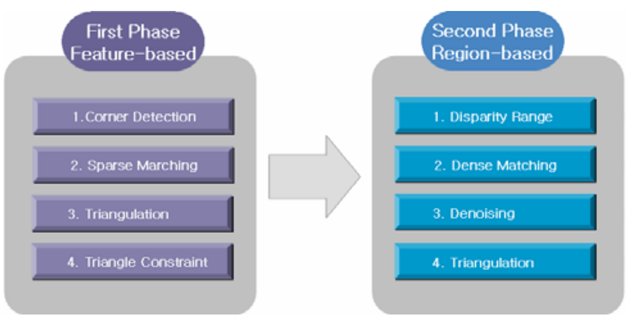


Fig. 1. Overview of the proposed stereo matching algorithm

First stage, triangle constraint is constructed through feature-based sparse matching. We take corners on the skin textures as feature, combined with image gradient, and epipolar constrains for sparse matching. After Delaunay triangulations, we found triangle constraint which is the preparation for dense matching.

Second stage, region-based dense matching is done after disparity formulation. In order to speed up matching process, we use a mask for each triangle region.

3.2 Theoretical Basis

This part refers to paper [18]. Suppose we have a continuous non self-occluding surface. To preserves the topology, each image must have the same topology as the surface does. The disparity gradient is defined as $\frac{|r_L - r_R|}{1/2|r_L + r_R|}$. Vector r_L is in left image, and its corresponding vector r_R is in right image. In the paper, there is a poof that a disparity gradient limits of less than 2 implies that the matches between the two images preserve the topology of the surface.

Suppose p_i, p_j are matched points, and q_i, q_j are another matched points. d_p and d_q are disparity of p and q respectively. $dis_{p_j q_j}$ is the distance between p and q in the second image. Illustration is shown in figure 2.

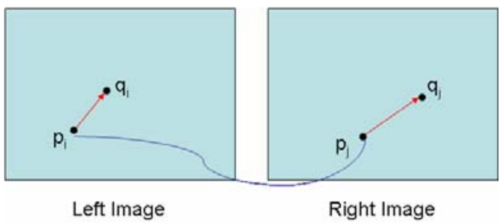


Fig. 2. Illustration of disparity gradient limits in which pixels p and q are in left image and right image

The disparity gradient is defined as follows:

$$G = \frac{|(p_i - q_i) - (p_j - q_j)|}{1/2|(p_i - q_i) + (p_j - q_j)|} = \frac{|(p_i - p_j) - (q_i - q_j)|}{1/2|(p_i - p_j) - (q_i - q_j) + 2(p_j - q_j)|} = \frac{|d_p - d_q|}{|1/2(d_p - d_q) + dis_{p_j q_j}|} \tag{1}$$

Because of the disparity gradient limit, let's suppose $G \le k < 2$ [17], then

$$|d_p - d_q| \le k|1/2(d_p - d_q) + dis_{p_j q_j}| \le k/2|d_p - d_q| + k \cdot dis_{p_j q_j} \tag{2}$$

Then, formulation (2) can be simplified

$$\left|d_p - d_q\right| \le \frac{2k}{2-k} dis_{p_j q_j} \tag{3}$$

So the disparity difference between two pixels in the image that taken from a non-occluding surface has the relationship with the distance of the two corresponding points. That means we can estimate the neighboring pixel's disparity through the known-disparity.

3.3 First Stage: Feature-Based Sparse Matching for Triangulation

The inputs to corner detection consist of a pair of rectified stereo images. In 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. Harris detector was used in the corner detection step and RANSAC algorithm to compute correspondence set to estimate the 2D homography and the (inlier) correspondences. We summarize the processing as follows:

1. Corner detection.
2. Choose window size w for correlation matching.
3. Look for the candidate matching points along the corresponding epipolar line.
4. Calculate the similarity [19] between the candidate matching points. Select the matched points.

With this strategy, only prominent corners in texture-rich areas are extracted. By parameter setting for Harris detector, we can control the density of triangulation. In order to improve the matching accuracy, we can change the searching area size. Parameter w is closer maximum disparity in global images, and accuracy is higher.

A Delaunay triangulation helps us to build the triangular map. It is similar to image segmentation as a prior for stereo matching algorithm. In our work, skin images are taken under narrow base line system considering light influence. Compared with the three triangle vertices' disparities in each region, disparity range belongs to $\left\{0, \max(V_1, V_2, V_3)\right\}$ for each to be matched triangle reasonably.

3.4 Second Stage: Region-Based Dense Matching

Along corresponding epipolar line, regional correlation stereo matching will be developed. We define an error energy function for region correlation criterion.

$$Error(d) = \frac{1}{9}\sum_{y=1}^{3}\sum_{c=1}^{3}(l(x, y+d, c) - r(x, y, c))^2 \tag{4}$$

Where, d is the disparity which is in disparity range from *0* to *Max*, c means the RGB color format and its value taken of *{1, 2, 3}*, and *l*, *r* represent the left image and right image. We have known the disparity of three vertexes in each matching area. Set the matched points as the seed. If the error is less than the threshold, associate the point into the region, and then choose the next points. If none points was selected, then give it up for next matching region.

4 Experiments

4.1 Skin Images Acquisition

The taken distance of test images is controlled in 15~20cm for clear skin pictures with surface texture details. In the close distance, long focus lens is necessary. For controlling two cameras to take photos at the same time, computer is necessary with camera SDK improvement. Our system is constituted by two Cannon 30D cameras with Tamron long focus lens 90mm. It performs on the personal computer with Intel Core 2Duo 2.66GHz CPU and 2G memory.

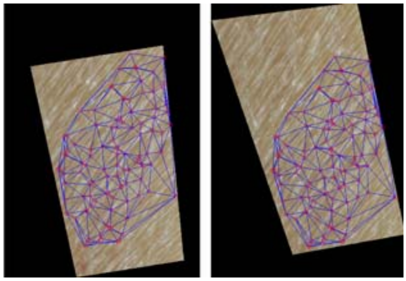
4.2 Experiment Results and Evaluation

The skin image pairs from human calf skin are shown in figure 3 with resolution 509*386. The real skin area is about 4mm*3mm. The biggest disparity value of pixel is 72. After sparse matching, triangulation is done in first stage. The result is shown in figure 4 with rectified images, in which red little circles on the two images represent the corresponding points after sparse matching, and number of matched points is 276. After we do triangulation, the triangle number is 201. Then disparity map is shown in figure 5. Image 5.a is result with our proposed algorithm, and image 5.b is the result with region based algorithm. Both of them are original data without filtering. There are many noises in 5.b, and disparity 5.a is the smoothing of 5.b. Total running time of matching is 102.70 second. The total number of matched points is 108,287. Suppose camera intrinsic matrix is 3×3 identity matrix, and we obtain 3D data points by self-calibration processing shown in figure 6 with a box corner and texture mapping effect.

Sparse matching result is related the system's precision. Table 1 lists the sparse matching results with the increase of detected corners. The accuracy of dense matching compared with manual measurement result is shown in table 2. We select sparse matching points according to figure 4.a manually. For dense matching, we select points in each triangle region randomly and find their matching points manually. Here number of 500 and 1000 points are considered respectively.

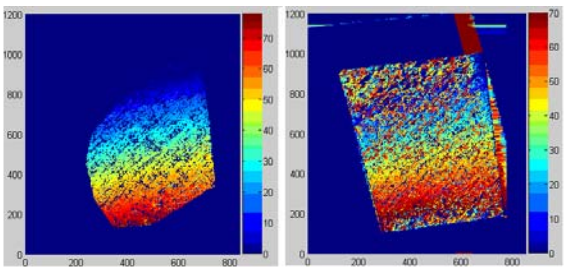
3.a 3.b

Fig. 3. Calf skin. Image 3.a is from left camera and image 3.b is from right camera.



4.a 4.b

Fig. 4. Feature-based matching and triangulation result



5.a 5.b

Fig. 5. Disparity map of calf skin images with proposed algorithm and region based algorithm respectively

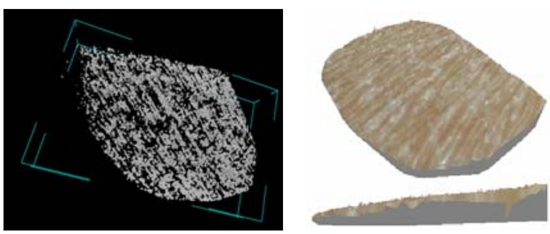


Fig. 6. 3D calf skin data and its visual effect after texture mapping

Table 1. The accuracy comparison of sparse matching in terms of pixel

Numberof corners	Number of matching points	Number of wrong matching points	Accuracy
886	171	0	100%
1340	276	0	100%
2476	480	5	98.96%

Table 2. The evaluation of the proposed algorithm in terms of pixel

Matching stage		Our algorithm		Region growing algorithm	
		Mean error	Max error	Mean error	Max error
Sparse matching: 276 matching points		0.48	1.25		
Dense matching	500 matching points	0.74	4.30	2.23	20.75
	1000 matching points	0.83	6.20	2.46	25.50

We show another stereo matching result in figure 7, in which pairs of image are taken from human thumb back skin. Image resolution is 516*468. The number of sparse matched points is 28. After we do triangulation, the triangle number is 43. Limited by the camera lens performance, some part of the thumb image is blurred. Our algorithm works in the clear region to keep the high accuracy.

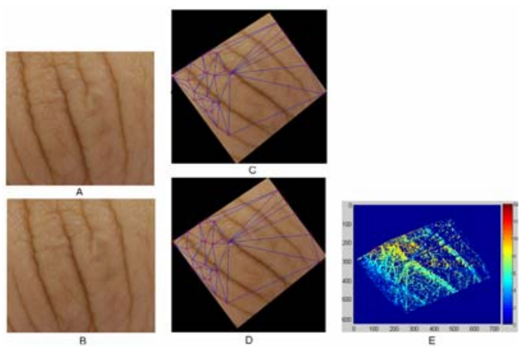


Fig. 7. Thumb skin Image. A, B on the left are original images. C, D in the middle are feature-based matching and triangulation result respectively. E is the finial disparity map.

5 Conclusion and Future Works

The human skin has the own specification properties, such as non-rigid object, micro skin texture structure. In the paper, we proposed a solution for micro skin surface reconstruction without 3D scanner. The solution for 3D reconstruction comes from the computer vision camera model and projective model. In order to reduce the light influence, we take images under narrow base line system. The two stages matching algorithm combined feature-based and region-based matching method has proposed for dense matching with higher accuracy. Considering disparity gradient limitation, we add the disparity constraints to each matching triangle region reasonably.

There are still some problems in our research. Skin wrinkle and texture with non-rigid property, we couldn't obtain the same skin state by time changing. For texture not well distributed skin, the triangular segmentation can be refined and output more matched points. Next step, to improve the performance, we will add more skin images for multi-view reconstruction.

Acknowledgement

This research was supported by the Ministry of Culture & Tourism and Korea Culture & Content Agency, Seoul, Korea, under Supporting Project of Culture Technology Institute.

References

1. Cula, O.-G., Dana, K.-J., Murphy, F.-P., Rao, B.-K.: Skin Texture Modeling. International Journal of Computer Vision 62(1-2), 97–119 (2005)
2. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, New York (2000)
3. Kolmogorov, V., Zabih, R.: Multi-camera Scene Reconstruction via Graph Cuts. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2352, pp. 82–96. Springer, Heidelberg (2002)
4. Veksler, O.: Stereo Correspondence by Dynamic Programming on a Tree. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 384–390. IEEE Press, New York (2005)
5. Wei, Y., Quan, L.: Region-Based Progressive Stereo Matching. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 106–113. IEEE Press, New York (2004)
6. Chiuso, A., Jin, H., Favaro, P., Soatto, S.: MFm: 3-D Motion and Structure from 2-D Motion Causally Integrated over Time Implementation. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1843, pp. 734–750. Springer, Heidelberg (2000)
7. Zhang, R., Tsai, P.-S., Cryer, J.-E., Shah, M.: Shape from Shading: a Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(8), 690–706 (1999)
8. Pavlidis, G., Koutsoudis, A., Arnaoutoglou, F., Tsioukas, V., Chamzas, C.: Methods for 3D Digitization of Cultural Heritage. Journal of Cultural Heritage (8), 93–98 (2007)
9. Pollefeys, M., Gool, L.-V., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual Modeling with a Hand-Held Camera. International Journal of Computer Vision 59(3), 207–232 (2004)
10. Michael, G., Noah, S., Brian, C., Hugues, H., Steven, M.-S.: Multi-View Stereo for Community Photo Collections. In: Proceedings of ICCV, pp. 1–8 (2007)
11. Pollefeys, M., Nister, D., Frahm, J.-M., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S.-J., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewenius, H., Yang, R., Welch, G., Towles, H.: Detailed Real-Time Urban 3D Reconstruction From Video. International Journal of Computer Vision 78(2), 143–167 (2008)
12. Kutulakos, K., Seitz, S.: A Theory of Shape by Space Carving. In: Proceedings of ICCV, pp. 307–314 (1999)

13. Faugeras, O., Keriven, R.: Variational Principles, Surface Evolution, PDE's, Level Set Methods and The Stereo Problem. IEEE Transactions on Image Processing 7(3), 336–344 (1998)
14. Fan, H., Ngan, K.-N.: Disparity Map Coding Based on Adaptive Triangular Surface Modeling. Signal Process.: Image Commun. 14(2), 119–130 (1998)
15. Lhuillier, M., Quan, L.: Matching Propagation for Image-Based Modeling and Rendering. IEEE Transactions on PAMI 27(3), 1140–1146 (2002)
16. Cech, J., Sara, R.: Efficient Sampling of Disparity Space for Fast and Accurate Matching. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
17. Pollefeys, M., Koch, R., Gool, L.-V.: A Simple and Efficient Rectification Method for General Motion. In: Proceedings of ICCV, pp. 496–501 (1999)
18. Trivedi, H.-P., Lloyd, S.-A.: The Role of Disparity Gradient in Stereo Vision. Perception 14(6), 685–690 (1985)
19. Aschwanden, P., Guggenbuhl, W.: Experimental Results from A Comparative Study on Correlation Type Registration Algorithms. In: Forstner, Ruwiedel (eds.) Robust computer vision, pp. 268–282 (1992)