Proc. VIIth Digital Image Computing: Techniques and Applications, Sun C., Talbot H., Ourselin S. and Adriaansen T. (Eds.), 10-12 Dec. 2003, Sydney

Fusion of Color Photometric Stereo Method and Slit Pattern Projection Method to Measure Three-Dimensional Shapes of a Human Face

Hitoshi Saji and Hiromasa Nakatani

Shizuoka University, Hamamatsu 432-8011, Japan

Abstract. We propose a new method for measuring three-dimensional (3D) moving facial shapes. This method uses photometric stereo method and slit pattern projection method with color light sources. The time required for sampling data can be reduced by switching the color light rays at high speed, and hence a series of 3D shapes of a moving human face at each instant can be measured. Experimental results are shown to demonstrate the feasibility of the proposed method.

1 Introduction

The human face is one of the most important parts of the human body because it has great expressive ability that can provide clues to the personality and the emotions of a person. Recently, automatic and non-contact facial analysis systems with video cameras and computers have been used in many applications. To analyze a human face correctly, the 3D shapes of the human face must be measured accurately.

In most of the existing methods, such as the binocular stereo method [1, 2], the 3D coordinates of objects are measured directly using two cameras. In such methods, the greatest difficulty is the identification of corresponding points of the images. The slit pattern projecting method simplifies the identification problem. By projecting points or lines onto a face, we can compute the 3D coordinates of the illuminated positions. To measure the entire face by the projecting method, the scanning method has been applied [3, 4, 5, 6]. This method, however, requires a lot of time to scan the entire face and the face should be kept motionless during the scan. Thus, the measuring time should be as short as possible to avoid errors due to changes in the shape of the face.

For rapid measurement, the use of a silicon range finder has been proposed [7]. This method can provide 3D coordinates at high speed. However, it requires special hardware for scanning the object, leading to a high production cost. Color-coded light stripes enable instantaneous measurement of the entire facial shape [8, 9]. The accuracy of the measured data, however, is low because it is affected by the color of the skin.

Recently, several methods of measuring the normal vectors of objects based on photometric properties, such as the shape from shading method [10, 11], and photometric stereo method [12, 13], have been proposed. These methods can provide a high density of 3D information about complicated object shapes in a short time. These methods, however, cannot be used to measure the 3D shapes of a human face accurately. We have integrated slit pattern projection method into photometric stereo method using a conventional video camera and measured the 3D shapes of a face accurately [14, 15]. However, this method uses only gray scale images and requires more than three images for the measurement.

In this paper, a new method of measuring a series of 3D facial shapes is introduced. This method uses a conventional video camera, two point light sources with red filters, and a slit pattern projector with a blue filter. The red light rays are projected onto a face at minute time intervals by switching the two point light sources, while the multiple blue light stripes are projected onto the face by the slit pattern projector all the time. We can measure the 3D facial shapes from the two images accurately by using the color information of each image. By switching the point light sources at high speed, the 3D shapes of a moving human face can be measured.

2 Method of Measurement

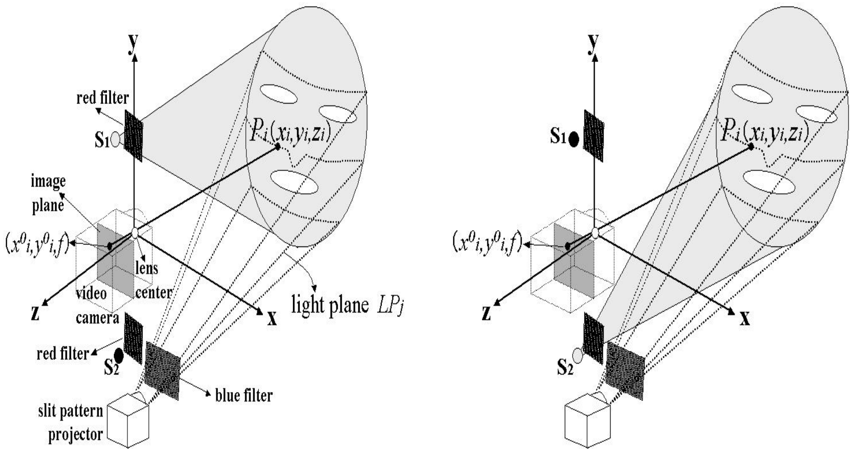
Fig. 1 shows the apparatus and the 3D orthogonal coordinate system for the proposed 3D measurement. A video camera is set in front of a face. The origin of the coordinate system is the center of the camera lens. The x-axis is parallel to the axis of the abscissa of the camera image plane and the y-axis is parallel to the axis of the ordinate of the plane. Two point light sources (S_1, S_2) and a slit pattern projector are arranged on a straight line that is parallel to the y-axis. Two red filters are set in front of the point light sources and one blue filter is set in front of the slit pattern projector.

With the apparatus, the proposed method is composed of five steps: (1) separation of facial images into two shading images and a slit image, (2) computation of normal vectors, (3) depth measurement by projecting stripes, (4) integration of normal vectors inside finite intervals, and (5) blending of curved segments.

2.1 Separation of Facial Images

Two point light sources S_1 and S_2 are switched in turn to illuminate the face, while a slit pattern projector illuminates the face all the time. The 3D facial shape at a particular instant is measured from two facial images. These images illuminated by each light source are obtained using the video camera. I_{S_1L} is the facial image illuminated by S_1 and the slit pattern projector. I_{S_2L} is the facial image illuminated by S_2 and the slit pattern projector. These images are color images described by

$$I_{S_1L} = (R_{S_1L}, G_{S_1L}, B_{S_1L}), \tag{1}$$
$$I_{S_2L} = (R_{S_2L}, G_{S_2L}, B_{S_2L}), \tag{2}$$



(a) (b)

Fig. 1. Two point light sources S_1 and S_2 are switched in turn to illuminate the face, while a slit pattern projector illuminates the face all the time. (a) The face is illuminated by S_1 and the slit pattern projector. (b) The face is illuminated by S_2 and the slit pattern projector.

where R_{S_kL}, G_{S_kL}, and B_{S_kL} are red component, green component, and blue component of I_{S_kL} (k = 1,2), respectively.

The face is illuminated by the point light and slit pattern at the same time. To distinguish between these light sources, we make sure that the range of the transmittance of the red/blue filter does not overlap with the range of blue/red sensitivity of the camera. Fig. 2 shows the ranges of the transmittance of the filters and the sensitivity of the camera in our system. The red component of the facial image corresponds to the intensity of point light sources S_1 and S_2, and the blue component of the facial image corresponds to the intensity of the slit pattern projector. With the apparatus, we separate the two facial images into two shading images I_{S_1} and I_{S_2} and a slit image I_L as follows:

$$I_{S_1} = R_{S_1L}, \quad I_{S_2} = R_{S_2L}, \quad I_L = B_{S_1L}. \tag{3}$$

I_{S_1} and I_{S_2} are used as the intensity of radiance in subsection 2.2. I_L is used as the slit image in subsection 2.3.

2.2 Computation of Normal Vectors

The normal vectors at all the points on the face are computed by the photometric stereo method [12]. In the conventional photometric stereo method, three light sources are necessary. In our system, only two light sources, S_1 and S_2, are used and they are switched at high speed to measure the vectors rapidly (Fig. 1). Two shading images of the face, I_{S_1} and I_{S_2}, illuminated by the corresponding light source, are obtained by the separation method described in subsection 2.1. From

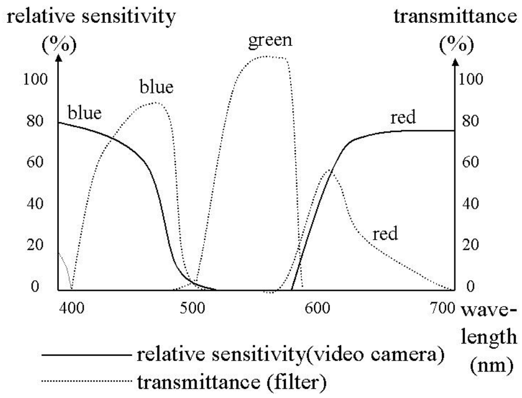


Fig. 2. Relative sensitivity of the video camera and transmittance of the red and blue filters.

the two images, one of the components (y component) of each normal vector on the face is determined by solving the Lambertian reflectance map. This method can provide a high density of information about the facial shape in a short time.

First, the normal vector $\mathbf{N}(x, y)$ at the point (x, y, z) on the face is described by

$$\mathbf{N}(x, y) = (N_x(x.y), N_y(x, y), 1). \tag{4}$$

Assuming that the light sources are far from the face, relative to the size of the face, the direction to each light source can be specified by a fixed vector. Two vectors, $\mathbf{S_1}$ and $\mathbf{S_2}$, which point in the direction from the point (x, y, z) to the light sources S_1 and S_2, respectively, are described by

$$\mathbf{S}_1 = (0, S_{1_y}, 1), \quad \mathbf{S}_2 = (0, S_{2_y}, 1). \tag{5}$$

Assuming the Lambertian reflectance map, the following two equations are given:

$$E_{S_1}(x, y) = e_{S_1} R \frac{\mathbf{S_1} \cdot \mathbf{N}(x, y)}{\|\mathbf{S_1}\| \|\mathbf{N(x, y)}\|}, \tag{6}$$

$$E_{S_2}(x, y) = e_{S_2} R \frac{\mathbf{S_2} \cdot \mathbf{N}(x, y)}{\|\mathbf{S_2}\| \|\mathbf{N(x, y)}\|}, \tag{7}$$

where e_{S_k} is the intensity of incident illumination from S_k filtering through the red filters, $E_{S_k}(x, y)$ is the intensity of radiance at the point (x, y) illuminated by $S_k(k = 1, 2)$, and R is the reflectance factor. The intensity e_{S_k} is set a priori and the intensity $E_{S_k}(x, y)$ is obtained from the pixel values of the two shading images $I_{S_k}(x, y)$ $(k = 1, 2)$.

From these equations, $N_y(x, y)$ is derived as

$$N_y(x, y) = \frac{1 - \alpha}{S_{2_y}\alpha - S_{1_y}}, \quad \alpha = \frac{E_{S_1}(x, y) e_{S_2} \|\mathbf{S_1}\|}{E_{S_2}(x, y) e_{S_1} \|\mathbf{S_2}\|} \quad . \tag{8}$$

2.3 Depth Measurement by Projecting Stripes

To obtain the boundary condition of integration, the proposed method uses the light projection method, i.e., multiple light stripes are projected onto the face using the slit pattern projector (Fig. 1) [16, 17]. The algorithm for the measurement is as follows:

1. Multiple light stripes are projected onto the face from one direction with the slit pattern projector.
2. The slit image I_L is obtained from the separation described in subsection 2.1.
3. Bright regions are extracted by thresholding the image.
4. The 3D coordinates of the points on the extracted bright regions are computed by the stereo vision algorithm.

Let $P_i = (x_i, y_i, z_i)$ $(i = 1, 2, 3, \cdots)$ be the points on the intersection of the face and the light plane LP_j. The coordinates (x_i, y_i, z_i) of P_i are computed from the following equations:

$$(l_{j_x}, 1.0, l_{j_z}) \cdot (x_i, y_i, z_i) = -d_j, \quad x_i = \frac{x_i^0}{f} z_i, \quad y_i = \frac{y_i^0}{f} z_i, \tag{9}$$

where f is the focal length of the camera, (x_i^0, y_i^0) are the coordinates of the intersection point of the image plane and the line passing through both the camera lens center and P_i, and Eq.(9) represents the plane LP_j.

2.4 Integration of Normal Vectors inside Finite Intervals

To reconstruct the 3D curved segment, we integrate the surface normals $\mathbf{N}(x, y)$ computed by the photometric stereo method in subsection 2.2. For the integration, the region of the face is projected onto the xy-plane and divided into small domains (Fig. 3).

Let P_i be the point on the light stripes on the face and the coordinates (x_i, y_i, z_i) of P_i be measured by the method described in subsection 2.3. Let the point $C_i = (x_i, y_i)$ be the projection of the measured point $P_i = (x_i, y_i, z_i)$ on the xy-plane. Fig. 4 shows the domain D_i of integration. The domain D_i is a two-dimensional finite interval centered at the point C_i on the xy-plane. The interval is parallel to the y-axis and the length of the interval is $2L_i$. To compute z coordinates at the point (x_i, y) in the domain D_i, that is $Z_i(x_i, y)$, the following equation is used:

$$Z_i(x, y) = \begin{cases} z_i - \int_{y_i}^{y} N_y(x_i, Y) dY \ (x = x_i), \\ \\ \qquad 0 \qquad\qquad (x \neq x_i). \end{cases} \tag{10}$$

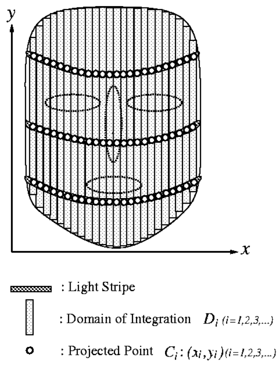


Fig. 3. Division of the surface region.

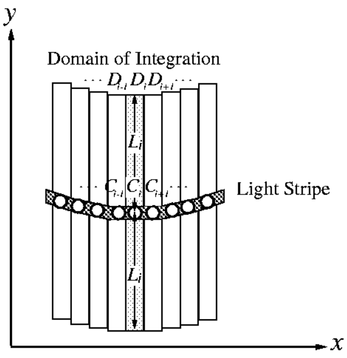


Fig. 4. The domain of integration.

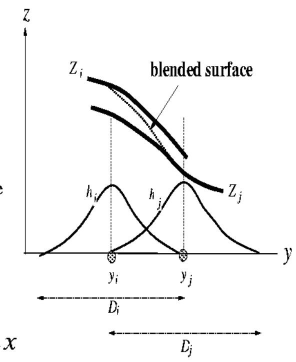


Fig. 5. Surface blending procedure.

2.5 Blending of Curved Segments

Since multiple light stripes are used, several projected points C_i have the same x coordinates. The integration domains are overlapped in several regions, such as D_i and D_j in Fig. 5, and the integrated values Z_i and Z_j, even on the same y coordinates, may not be the same in the different domains. Therefore, computed curved segments in the overlapped regions of the domains are blended using infinitely differentiable functions.

In each domain D_i, the blending function h_i is given as follows [18]:

$$h_i(x,y) = \begin{cases} a\left(\frac{L_i - |y-y_i|}{L_i}\right) & (if \quad |y-y_i| < L_i \ and \ x = x_i) \\ 0 & (otherwise) \end{cases}, \tag{11}$$

where $a(t) = \begin{cases} 0 & (t \le 0) \\ \frac{b(t)}{b(t)+b(1-t)} & (0 < t < 1) \\ 1 & (1 \le t) \end{cases}$ and $b(t) = \begin{cases} 0 & (t \le 0) \\ e^{-\frac{1}{t}} & (0 < t) \end{cases}$.

Clearly, the function $h_i(x,y)$ is infinitely differentiable. Now, the surface $Z(x,y)$ is computed as follows:

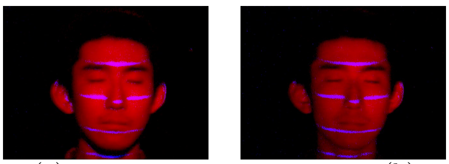
$$Z(x,y) = \frac{\sum_i h_i(x,y) Z_i(x,y)}{\sum_i h_i(x,y)}, \tag{12}$$

where $Z_i(x,y)$ are the z coordinates of the segment at (x,y) in the domain D_i.

3 Experiments

Fig. 6 shows the experimental apparatus and Fig. 7 shows two facial images taken by the conventional video camera. In these images, a subject is illuminated by the light sources S_1, S_2, and the slit pattern projector. The distance between the light sources and the subject face is about 3.0 meters. All the light rays are in the visible spectrum range and they do not have any adverse effects on human beings. The size of the two facial images is 640×480 pixels. Fig. 8 shows the separated three images, i.e. two shading images and one slit image. Fig. 9 shows the 3D facial shape obtained from the separated images. The two light sources are switched on in turn at a rate of 15 times per second. The facial images of the subject are taken at a rate of 15 frames per second using the video camera. Therefore, a series of 3D facial shapes at 15 frames per second can be measured. The measuring time can be reduced by using a high-speed video camera and by synchronizing image capturing with light switching. The computation time for obtaining a facial shape at an instant is about 10 seconds in total on a personal computer with a 1GHz cpu. This time can be reduced by using a parallel processing method for computing normal vectors and integrating the vectors. As shown in Fig. 10, the time sequence of the 3D facial shapes is obtained from the time sequence of the images. Fig. 11 shows the selected time sequence of the measured 3D facial shapes.

Fig. 6. Experimental apparatus



(a) (b)

Fig. 7. (a) Facial image $I_{S_1 L}$ (b) Facial image $I_{S_2 L}$

(a) (b) (c)

Fig. 8. (a) Shading image I_{S_1} (b) Shading image I_{S_2} (c) Slit image I_L

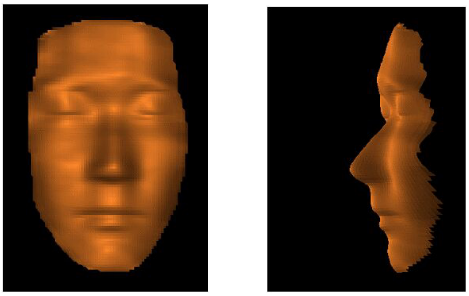


Fig. 9. 3D facial shape.

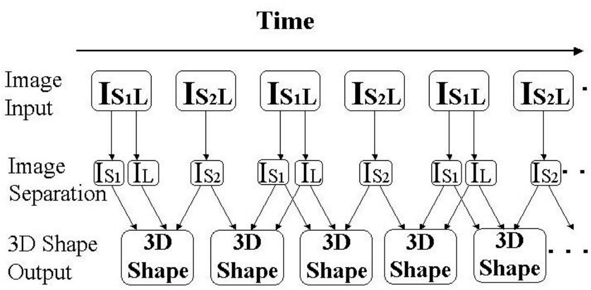


Fig. 10. Time sequence of the measurement process.

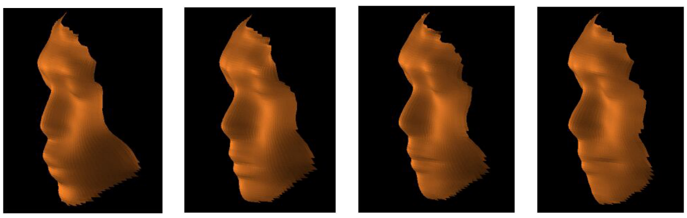


Fig. 11. Time sequence of 3D facial shapes.

4 Evaluation of the Proposed Method

The proposed method is evaluated by measuring the 3D facial shapes of 20 motionless subjects using both the proposed method and another accurate method simultaneously and comparing the coordinates measured by the two methods. To obtain accurate 3D coordinates, the binocular stereo method is employed, and the corresponding points in the two images taken by the two cameras are identified manually. The z coordinates on the bridge of the nose where the shape is most confusing have the largest margin of error.

The maximum error of 20 subjects in the regions between two light stripes illuminated by the slit projector is smaller than 3mm, while the maximum error on the upper end (head) or lower end (chin) of the face is smaller than 7mm.

5 Summary and Conclusions

In this paper, we introduced a new method for measuring 3D shapes of a moving human face. This method uses two point light sources with red filters, a slit pattern projector with a blue filter, and a conventional video camera. The point light sources are switched in turn to illuminate the face, while the slit pattern projector illuminates the face all the time. All the light rays are in the visible spectrum range and they do not have any adverse effects on human beings. With the apparatus, a series of 3D facial shapes can be measured at 15 frames per second. Even if the human face moves and changes, detailed 3D facial shapes at a particular instant can be measured accurately.

The 3D deformations of the facial shapes are obtained from the time sequence of the measured 3D facial shapes. Moreover, the time sequence of the texture on the face is also obtained from the intensity of the facial image. The measuring time can be reduced by using a high-speed video camera and by synchronizing image capturing with light switching. As a further possible improvement of the system, the use of infrared light sources to avoid the influence of ambient light on facial expressions is now being considered.

There are two main targets for future research in the field of human facial analysis. One is the identification of individual human faces and the other is the analysis of human facial expressions. In the former, conventional methods have been applied to obtain facial features from a stationary face. However, if the face moves, the person cannot be identified by the conventional methods. Using the proposed method, 3D shapes are measured from a moving human face and several facial features can be obtained easily from the measured 3D data. Hence, the proposed method can be used to identify a person correctly. In the latter, conventional methods are used to analyze human facial expressions by obtaining a few features from the moving face, such as the shapes of the eyes, eyebrows, and mouth. To analyze subtle human facial expressions, it is necessary to obtain detailed shapes, such as wrinkles on the cheeks and forehead. The proposed method can be used to measure detailed 3D shapes over the complete facial region and can be applied to the analysis of subtle changes in human facial expressions.

References

1. S. T. Barnard and M. A. Fischler. "Computational Stereo". *ACM Computing Surveys*, 14(4):553–572, December 1982.
2. B. D. F. Methley. *Computational Models in Surveying and Photogrammetry.* Blackie & Son Ltd, 1986.
3. M. Nahas, H. Huitric, M. Rioux, and J. Domey. "Registered 3D-Texture Imaging". *Computer Animation '90*, pages 81–91, 1990.
4. Y. Watanabe and Y. Suenaga. "Synchronized Acquisition of Three-Dimensional Range and Color Data and its Applications". *Scientific Visualization of Physical Phenomena*, pages 655–665, 1991.
5. M. W. Vannier, T. Pilgram, G. Bhatia, and B. Brunsden. "Facial Surface Scanner". *IEEE Computer Graphics and Applications*, 11(6):72–80, November 1991.
6. F. Chen, G.M.Brown, and M.Song. "Overview of three-dimensional shape measurement using optical methods". *Opt. Eng.*, 39(1):10–22, 2000.
7. K. Sato. "Silicon Range Finder –A Realtime Range Finding VLSI Sensor–". *IEEE 1994 Custom Integrated Circuits Conference*, pages 339–342, 1994.
8. T. Abe and M. Kimura. "A Measurement Method of 3-D Shape of the Human Face Using Color-Coded Light Stripes". *The Transactions of the Institute of Electronics and Communication Engineers of Japan*, J72-DII(12):2061–2069, 1989.
9. Z. J. Geng. "Rainbow three-dimensional camera: new concept of high-speed three-dimensional vision systems". *Opt. Eng.*, 35(2):376–383, 1996.
10. B. K. P. Horn and M. J. Brooks. "The Variational Approach to Shape from Shading". *Computer Vision, Graphics, and Image Processing*, 33(2):174–208, 1986.
11. B. K. P. Horn and M. J. Brooks, editors. *Shape from Shading.* The MIT Press, Cambridge Massachusetts London, 1989.
12. R. J. Woodham. "Photometric Method for Determining Surface Orientation from Multiple Images". *Opt. Eng.*, 19(1):139–144, 1980.
13. H. Saji, H. Hioki, Y. Shinagawa, K. Yoshida, and T. L. Kunii. "Extraction of 3D Shapes from the Moving Human Face Using Lighting Switch Photometry". *Computer Animation 92*, pages 69–86, 1992.
14. H.Saji. "Measuring Three-Dimensional Shapes of a Human Face Using Photometric Stereo Method with Two Light Sources and Slit Patterns". *Proceedings of MVA'96 IAPR Workshop on Machine Vision Applications*, pages 297–300, 1996.
15. H.Saji and H. Nakatani. "Measuring Three-Dimensional Shapes of a Moving Human Face Using Photometric Stereo Method with Two Light Sources and Slit Patterns". *IEICE Transactions on Information and Systems*, E80-D(8):795–801, 1997.
16. R. A. Jarvis. "A Perspective on Range Finding Techniques for Computer Vision". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(2):122–139, 1983.
17. F. Röcker and A. Kiessling. "Methods for Analyzing Three Dimensional Scenes". *4th. International Joint Conference on Artificial Intelligence*, pages 669–673, 1975.
18. S. Takahashi and T. L. Kunii. "Manifold-based Multiple Viewpoint CAD : A Case Study of Mountain Guide Map Generation ". *Computer Aided Design*, 26(8):622–631, August 1994.