

# 3D human face reconstruction based on band-limited binary patterns

Xin Fan (范鑫)<sup>1,2,3</sup>, Changhe Zhou (周常河)<sup>1,\*</sup>, Shaoqing Wang (王少卿)<sup>1</sup>,  
Chao Li (李超)<sup>3</sup>, and Boquan Yang (杨博荃)<sup>4</sup>

<sup>1</sup>Laboratory of Information Optics and Optoelectronics Techniques, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Science, Shanghai 201800, China

<sup>2</sup>University of Chinese Academy of Science, Beijing 100049, China

<sup>3</sup>School of Physical Science and Technology, ShanghaiTech University, Shanghai 200031, China

<sup>4</sup>College of Science, Shanghai University, Shanghai 201800, China

\*Corresponding author: chazhou@mail.shcnc.ac.cn

Received April 13, 2016; accepted May 24, 2016; posted online June 21, 2016

Face recognition technology has great prospects for practical applications. Three-dimensional (3D) human faces are becoming more and more important in consideration of the limits of two-dimensional face recognition. We propose an active binocular setup to obtain a 3D colorful human face using the band-limited binary patterns (BBLP) method. Two grayscale cameras capture the BBLP projected onto the target of human face by a digital light processing (DLP) projector synchronously. Then, a color camera captures a colorful image of the human face. The benefit of this system is that the 3D colorful human face can be obtained easily with an improved temporal correlation algorithm and the precalibration results between three cameras. The experimental results demonstrated the robustness, easy operation, and the high speed of this 3D imaging setup.

OCIS codes: 110.6880, 150.6910, 330.1400, 330.1720.

doi: 10.3788/COL201614.081101.

Face recognition technology has great prospects in public security, financial and homeland security, etc.<sup>[1,2]</sup>. However, face recognition is confronted with some challenges that prevent its widespread use due to its low recognition accuracy in essential two-dimensional (2D) representations. Traditional recognition methods are mostly based on 2D photographs, which are easily confused because of their different head pose, lighting difference, facial expressions, and other characteristics. In order to overcome these difficulties, more and more attention has been paid to three-dimensional (3D) human face acquisition technique, because a 3D colorful human face model can offer more information, just like human eyes<sup>[3,4]</sup>. However, 3D face reconstruction brings new challenges because the human faces are lowly textured and it is hard to obtain an accurate 3D model. Some commercial laser scanners have been employed to directly capture 3D face data, but the high expense of this equipment and the low speed of reconstruction make them difficult to popularize<sup>[5]</sup>. Methods based on prior face models are proposed. Kemelmacher and Basri proposed a single-image reconstruction method using a shape-from-shading approach that requires only a single template face as a prior<sup>[6]</sup>. This approach can yield good results, but the geometry varies significantly depending on which image and template are used. Several methods based on stereo passive vision were also proposed<sup>[4]</sup>. However these methods are sensitive to the lighting conditions.

In this Letter, a non-contact active scanner<sup>[7,8]</sup> for a 3D colorful human face is proposed. A color camera is attached to standard binocular cameras to obtain the color texture information of a human face. An optimal temporal

correlation technology (OTCT) is also proposed to improve the accuracy of the corresponding points.

Figure 1 shows a schematic view of the scanner. The up and down objects are two fire-wire cameras with 2.0 MP (1600 × 1200, pixel size = 4.5 μm). The focal length of the camera lenses is 16 mm, and the baseline of the two cameras is about 30 cm. A digital light processing (DLP) projector with a focal length of 80 cm is placed between these two cameras to project special patterns onto a human face, and a color camera with 22 MP (5760 × 3840, pixel size = 6.25 μm) is near the DLP projector. A series

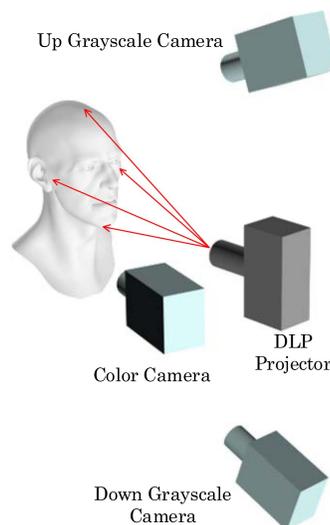


Fig. 1. Schematic view of the active binocular 3D setup.

of binarized band-limited patterns with a resolution of  $912 \times 1140$  are projected onto the human face via the DLP projector during the measurement. All the cameras were calibrated with a MATLAB toolbox before the experiments<sup>[9]</sup>.

Finding homologous points is a key step for all 3D reconstruction using stereo vision. Traditional matching methods, such as the sum of the squared difference, sum of absolute intensity value differences, and normalized cross correlation, simply perform template matching in the spatial domain. All of these methods are based on the texture features of objects and the results mostly depend on the quality of the images, which are easily affected by the environmental light and camera angle. These methods performed badly when reconstructing a smooth surface. Davis *et al.* proposed a method named temporal correlation technology (TCT) to extend the length of the correlation windows to pixels in the temporal domain<sup>[10]</sup>. For the TCT-based stereo-matching problem, the result is given by

$$\begin{aligned} \text{TCT}(x, y, d) \\ = \frac{\sum_{t=1}^N [I_U(x, y, t) - M_U] \cdot [I_D(x + d, y, t) - M_D]}{S_U(x, y, t) \cdot S_D(x + d, y, t)}, \quad (1) \end{aligned}$$

where the numerator of Eq. (1) represents the cross correlation between the two temporal intensity values,  $I_U$  and  $I_D$  represent the intensities of the binarization images, and  $d$  represents the disparity between the up and down images. The search window is only along the horizontal direction because of epipolar rectifying.  $M_i$  and  $S_i$  denote the mean intensity value and the standard deviation of the temporal intensity vector in the up and down images, i.e.,

$$M = \frac{1}{N} \sum_{t=1}^N I(u, v, t), \quad (2)$$

$$S = \sqrt{\frac{1}{N} \sum_{t=1}^N (I(u, v, t) - M)^2}, \quad (3)$$

and two pixels can be homologous when the TCT exceeds the threshold and reaches a maximum.

Liu *et al.*<sup>[11]</sup> have proposed a setup to measure the ground surface of an optical element based on the TCT method, and it performs well when measuring low-textured glass just like a human face. However, because the edge pixels of the patterns may be lost when the images are processed into binary, traditional TCT is easily leads to wrong matching. We expand the matching window to the neighborhood of matching pixels to combine the temporal and spatial domains and propose a method named OTCT. The OTCT is given by

$$\text{OTCT}(x, y, d) = \frac{\sum_{t=1}^N \sum_{dx=-m}^m \sum_{dy=-n}^n [I_U(x + dx, y + dy, t) - M_U] \cdot [I_D(x + dx + d, y + dy, t) - M_D]}{S_U(x, y, t) \cdot S_D(x + d, y, t)}, \quad (4)$$

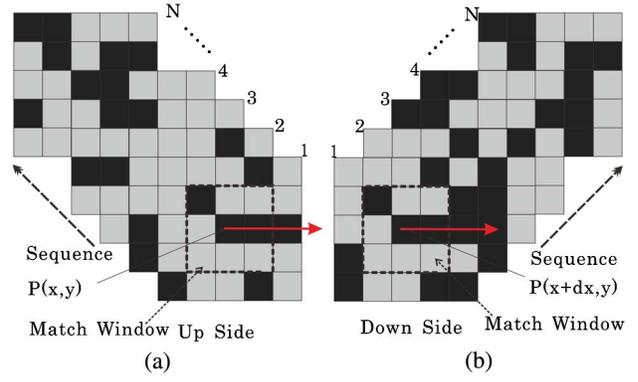


Fig. 2. Homologous points matching window using OTCT.

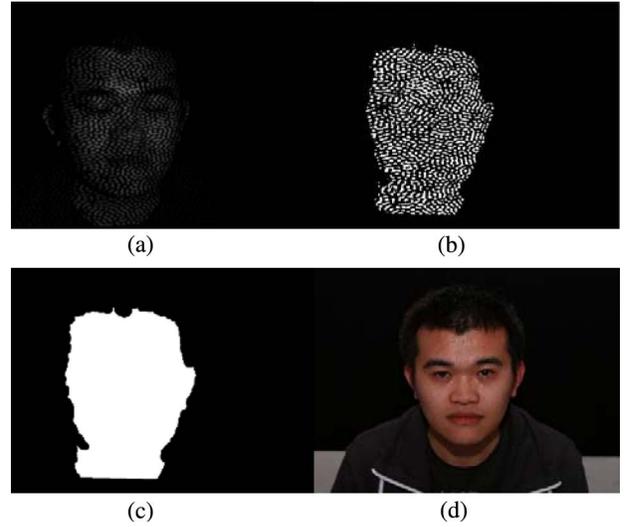


Fig. 3. (a) and (b) One rectified image in experiment and the result of self-adapting binarization. (c) Extraction of the region of interest.

where  $dx$  and  $dy$  represent the neighboring pixels of the matching point in two directions. Obviously,  $m$  and  $n$  represent the matching window's size; the matching window is  $(2m + 1) \times (2n + 1)$ , as shown in Fig. 2.

During the measurement, a series of  $N$  binary band-limited random patterns are projected onto the human face, and two grayscale cameras will capture images synchronously. After pattern sequence, one color image is captured by the color camera, as shown in Fig. 3(d). The rectified images can be obtained using the precalibration results. After extracting the region of interest, self-adapting binarization, and using the triangle principle<sup>[12]</sup>, the dense 3D point cloud of the human face will be reconstructed. By re-projecting the point cloud back to the

**Table 1.** Comparison Between Different Pattern Numbers Using OTCT with the Matching Window of  $3 \times 3$

Pattern number (N)	Accurate points ( $10^4$ )
10	13.7
15	19.2
20	20.0

**Table 2.** Comparison Between Different Matching Windows with Pattern Number of 20

Matching method	Matching window	Accurate points ( $10^4$ )	Time (s)
TCT	$1 \times 1$	11.6	401
OTCT	$3 \times 3$	20.0	1253
OTCT	$5 \times 5$	20.1	3705

color image, the texture information of the human face can be obtained easily.

Table 1 shows the accuracy comparison of OTCT with a matching window of  $3 \times 3$  when using different pattern numbers to obtain a human face. Our program is run by a personal computer with a CPU core of a 2.6 GHz clock frequency under Matlab 2012. The amount of accurate points increases as the pattern numbers projected onto the human face increases. Table 2 shows the reconstruction accuracy and time comparison between TCT and OTCT when the pattern number is 20. The computation time increases when using a bigger matching window. The suggested matching window is  $3 \times 3$  based on this table, which will be helpful to obtain better results with less calculating time. Figures 4(a) and 4(b) show the comparison between TCT and OTCT. It is obvious our method can get better results when using the same pattern number. Figure 4(b) shows the reconstruction results of the human face with OTCT. The related point cloud consists of more than  $2.0 \times 10^5$  points and the absolute error of the full field is less than 0.4 mm.

In conclusion, we develop an active binocular 3D setup based on band-limited patterns to obtain a colorful human face. We expand TCT to the spatial domain to improve the accuracy of the corresponding points. The experiments show that our OTCT method performs better when using fewer patterns. And the suggested matching window is

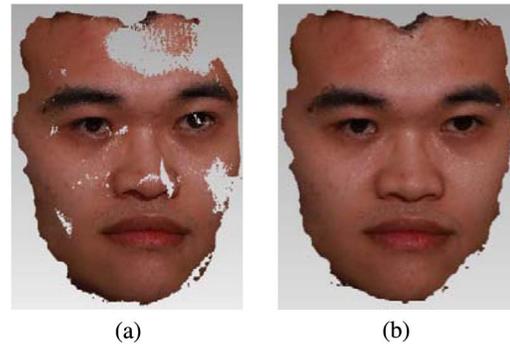


Fig. 4. Reconstruction results of human face with (a) TCT and (b) OTCT methods.

$3 \times 3$ , which can obtain better results with a low computational complexity. With the color camera, the texture information of the human face can be obtained easily. The experimental results verify the robustness, easy operation, and the high speed of this method.

This work was supported by the National Natural Science Foundation of China (No. 61308073) and the Science and Technology Commission of Shanghai Municipality (No. 15JC1403500).

## References

1. W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, *ACM Comp. Sur.* **35**, 399 (2003).
2. C. Zhou, "Three-dimensional identification card and related methods," Chinese Patent CN201210253349.7 (2012).
3. J. Kittler, A. Hilton, M. Hamouz, and J. Illingworth, in *IEEE Compute Society Conference on Computer Vision and Pattern Recognition* 114 (2005).
4. D. M. Levine and Y. C. Yu, *Pattern Recognit. Lett.* **30**, 908 (2009).
5. Y. Zheng, J. Chang, Z. Zheng, and Z. Wang, in *IEEE International Conference on Image Processing 3* (2007).
6. K. Shlizerman Ira and R. Basri, *IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 394 (2010).
7. F. Chen, G. M. Brown, and M. Song, *Opt. Eng.* **39**, 10 (2000).
8. Z. Zhao, J. Wu, Y. Su, N. Liang, and H. Duan, *Chin. Opt. Lett.* **12**, 091101 (2014).
9. J. Y. Bouguet, "Camera calibration toolbox for Matlab," [http://www.vision.caltech.edu/bouguetj/calib\\_doc/index.html](http://www.vision.caltech.edu/bouguetj/calib_doc/index.html) (2004).
10. J. Davis, R. Ramamoorthi, and S. Rusinkiewicz, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2**, 359 (2003).
11. K. Liu, C. Zhou, S. Wang, S. Wei, and X. Fan, *Chin. Opt. Lett.* **13**, 081101 (2015).
12. K. Liu, C. Zhou, S. Wei, S. Wang, X. Fan, and J. Ma, *Appl. Opt.* **53**, 6083 (2014).