A Mobile Structured Light System for 3D Face Acquisition

Marco Piccirilli, *Student Member, IEEE*, Gianfranco Doretto, *Member, IEEE*,
Arun Ross, *Senior Member, IEEE*, and Donald Adjeroh, *Member, IEEE*

***Abstract*— A mobile sensor based on fringe projection techniques is developed with the goal of acquiring face 3D and color with a smartphone device. The system consists of a portable pico-projector and an Android-based smartphone. The data acquisition, pattern generation. and reconstruction of the final 3D point cloud are all driven by the smartphone. We present results on the root-mean-square error (RMSE) of the sensor and on 3D face matching.**

***Index Terms*— Depth sensor, structured light, mobile device, 3D face acquisition, 3D face matching, 3D biometrics.**

I. Introduction

THE increased availability of handheld optical 3D scanning devices as well as the maturing of 3D printing options have led to a new interest in 3D scanners in general. Although Kinect[1] type of scanners can cheaply acquire a stream of depth images, they are not built for mobile devices. Recent scanners like the Occipital Structure,[2] and the Project Tango[3] are tailored to generic mobile indoor 3D modeling applications, rather than acquiring face 3D data in uncontrolled scenarios, where the position and illumination of the imaged subjects are not predefined. Our *goal*, instead, is to build a mobile depth and color acquisition system on a smartphone, offering good accuracy, and short capture time for acquiring face data in uncontrolled conditions.

3D acquisition using passive stereo vision heavily depends on the nature of the surfaces being imaged, and on the presence of surface texture. Active illumination techniques remove this dependence, speeding up the reconstruction step, and increasing resilience against illumination variations. However, our specific goal does not allow to simply embed an off-the-shelf algorithm. Indeed, the main contribution of this work is a unique integration of existing methods, driven by the hardware limitations and by the constraints of our application, which enables a new 3D face acquisition approach in unconstrained scenarios based on a smartphone.

A minimal active system is composed of a light source and a camera. In [1], the smartphone screen is used as a light source, but its power drastically limits the acquisition distance and accuracy. Instead, our system uses a pico-projector as a portable structured light source, and can be used indoors and outdoors. This work constitutes the first time a pico-projector is being driven by a smartphone for 3D face acquisition.

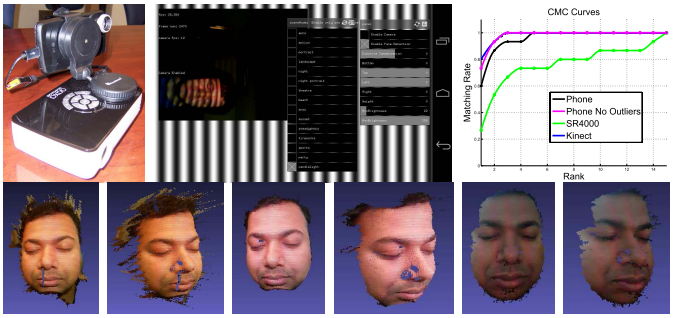

Fig. 1. System setup (top-left), and acquisition application (top-center). Face recognition CMC curves (top-right). Face depth and texture maps with outliers (two bottom-left images), with outliers filtered out (two bottom-center images), and without outliers captured outdoors (two bottom-right images).

II. Methods

Our sensor system is composed by a LG Nexus 4 smartphone, driving an AAXA P300 pico-projector. The smartphone acquires images at a resolution of 640 × 480 pixels. The projector illuminates the scene with fringe patterns having resolution of 1024 × 968 pixels, emitting 300 lumens. The smartphone mounts an additional 2× lens for pairing its field of view with the projector field of view (see Fig. 1 (top-left)). The system battery life is limited by the projector, which is one hour at maximum power. The smartphone runs Android OS, and three mobile applications we implemented based on the Android SDK, NDK, and OpenFramework. The first one performs data acquisition, the second computes the 3D reconstruction of a face, and the third is used for the calibration of the system. The following sections provide details about, and motivations for the algorithms we integrated in these applications.

1) Unconstrained Scenarios: In fringe projection methods, the scene is illuminated by fringe patterns and several images are captured. Measures of fringe distortions are used to recover depth. While patterns are projected, the subject being captured is expected to remain still, which is not the case in unconstrained scenarios. Thus, minimizing capture time is essential for accurate acquisitions. This requires the smallest number of patterns. While Gray-codes based approaches use tens of patterns for one depth map estimation, phase-shifting methods reduce them by a 10-fold factor, and are very robust [2].

Traditional phase-shifting strategies, [2], estimate depth from phase increments, computed with respect to a reference plane. This entails taking two acquisitions without moving

Manuscript received June 12, 2015; accepted December 8, 2015. Date of publication December 22, 2015; date of current version February 10, 2016. This work was supported by the National Science Foundation Office within the Director Industry and University Cooperative Research Program, Center for Identification Technology Research under Grant 1066197. The associate editor coordinating the review of this letter and approving it for publication was Prof. Kazuaki Sawada.

M. Piccirilli, G. Doretto, and D. Adjeroh are with the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506 USA (e-mail: mpiccir1@mix.wvu.edu; gianfranco.doretto@mail.wvu.edu; donald.adjeroh@mail.wvu.edu).

A. Ross is with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824 USA (e-mail: rossarun@cse.msu.edu).

Digital Object Identifier 10.1109/JSEN.2015.2511064

[1]http://www.microsoft.com/en-us/kinectforwindows/

[2]http://structure.io

[3]http://www.google.com/atap/project-tango

the sensor. The first one with only the reference plane, and the second with the subject in front of it. However, this approach cannot be used for face capture in unconstrained scenarios, from noncooperative subjects and a moving sensor. The only techniques that do not require a reference are the stereo fringe projection approaches, where depth estimation is based on triangulation methods. We exploit the *camera-projector epipolar (CPE)* algorithm [3] for computing depth. Compared to others, CPE requires the generation of only one sequence of fringe patterns, further shortening the sensor acquisition time by a factor of 2. CPE requires the intrinsic, extrinsic, and lens distortion calibration parameters, which we estimate with the third mobile application according to [4].

2) Phase-Shifting Algorithm: We use a 3-step phase-shifting method [2], with a 16-fringe pattern, and a shift of $2\pi/3$, to encode and decode phase information. This provides a number of advantages, including an excellent trade-off between speed and robustness. It is robust to noise and illumination variation, allowing sensor use in indoor and outdoor conditions. Moreover, a sequence of fringe patterns is captured by three images, as opposed to methods that require more. This reduces the acquisition time, and increases robustness against the relative motion between sensor and imaged subject. Our system projects a fringe pattern and acquires an image every 0.05 s, leading to a joint color and depth map acquisition every 0.143 s. The 3-step approach has also the advantage of recovering the texture map without acquiring an extra image.

The acquisition application (Fig. 1(top-middle)), drives the pico-projector and allows setting, among other parameters, the fringe period, and the min and max intensities to avoid image saturation. The 3D reconstruction application includes a fast implementation of the 3-step phase-shifting decoding for recovering the wrapped phase [2]. This algorithm is very light, based on a lookup table, and runs in real time on state-of-the-art smartphones. A gamma correction of the images is introduced to compensate for the non-linearities of the projector and the camera.

3) Phase Unwrapping: The wrapped phase exhibits sawtooth-like discontinuities of 2π. To obtain a continuous phase map, those are removed by an unwrapping method. To minimize the acquisition time, we deploy spatial unwrapping, as opposed to temporal unwrapping, which requires the acquisition of more frames. We use the multilevel quality-based unwrapping [5]. It computes a quality map for guiding the unwrapping of the phase on the pixels with highest quality, making the process faster and more reliable.

4) 3D Face Reconstruction: We improve the acquisition by integrating the smartphone built-in face detector. In particular, we tune the camera metering and focus to the face area of the imaged subject, leading to a well exposed and in-focus acquisition. Moreover, we use the face bounding box information to select the pixels for the 3D face reconstruction algorithm. Specifically, the phase and position of pixels inside the box and with quality (provided by the phase unwrapping [5]) above a certain threshold are used by the CPE algorithm to recover depth and texture. This approach limits the number of points to be processed (usually around $50,000$), further improving the speed of the reconstruction. The two bottom-left images of Fig. 1 show the results of this approach. Notice that background pixels with high quality can become part of the face even if their depth is significantly different, leading to visible artifacts. This is due to the inability of the spatial unwrapping to correctly handle depth discontinuities bigger than 2π. We address this issue by computing the PCA of the points within a ball of radius 1/4 of the smallest edge of the bounding box, and centered on the centroid. The third principal component estimates the head orientation with respect to the sensor. Points with depth, projected onto this component, outside of a prefixed range are filtered out as outliers. The two bottom-center images of Fig. 1 show the improved result using this method, and the two bottom-right images show an outdoor acquisition.

III. Results

We tested the system for mobile 3D face identification. We implemented a face matcher based on viewpoint feature histogram descriptors [6], compared with the χ^2-distance. Fig. 1(top-right) shows the cumulative match characteristic curve (CMC) for our system and two other devices: the Kinect-V1 and the Mesa SR4000 TOF camera, all placed at 1 m from the face of the subjects. Our sensor provides results comparable to those from the Kinect-V1, and largely outperforms the TOF camera, also due to its lower resolution (176×144 pixels).

We compared the face depth data acquired with our sensor, the Kinect-V1, and the TOF camera against the depth maps acquired with a Minolta Vivid 910 3D laser scanner with tele lens, and Z accuracy of ± 0.10 mm. After registering the face point clouds, the RMSE is 4.64 mm for our device, 4.15 mm for the Kinect-V1, and 16.66 mm for the TOF camera.

IV. Conclusion

We developed a portable 3D face acquisition system based on a smartphone and a pico-projector. The system leads to results comparable to those from the Microsoft Kinect for XBox 360. However, the proposed system is portable, battery-powered, and can work outdoors.

References

[1] G. Schindler, "Photometric stereo via computer screen lighting for real-time surface reconstruction," in *Proc. 4th Int. Symp. 3DPVT*, 2008, pp. 1–6.

[2] P. S. Huang and S. Zhang, "Fast three-step phase-shifting algorithm," *Appl. Opt.*, vol. 45, no. 21, pp. 5086–5091, Jul. 2006.

[3] C. Bräuer-Burchardt, M. Möller, C. Munkelt, M. Heinze, P. Kühmstedt, and G. Notni, "On the accuracy of point correspondence methods in three-dimensional measurement systems using fringe projection," *Opt. Eng.*, vol. 52, no. 6, p. 063601, 2013.

[4] D. Moreno and G. Taubin, "Simple, accurate, and robust projector-camera calibration," in *Proc. 2nd Int. Symp. 3DIMPVT*, Oct. 2012, pp. 464–471.

[5] S. Zhang, X. Li, and S.-T. Yau, "Multilevel quality-guided phase unwrapping algorithm for real-time three-dimensional shape reconstruction," *Appl. Opt.*, vol. 46, no. 1, pp. 50–57, Jan. 2007.

[6] R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu, "Fast 3D recognition and pose using the viewpoint feature histogram," in *Proc. IEEE/RSJ Int. Conf. IROS*, Oct. 2010, pp. 2155–2162.