Using ToF camera and two mirrors for 3D reconstruction of dynamic objects

Technical Report Number 1380

Trong-Nguyen Nguyen*[1], Huu-Hung Huynh†[2], and Jean Meunier‡[1]

[1]DIRO, University of Montreal, Montreal, QC, Canada
[2]University of Science and Technology, Danang, Vietnam

September 8, 2018

Abstract

Vision systems are currently playing an important role in a wide range of applications. Many recent systems attempted to represent the object of interest as a 3D model instead of a sequence of 2D images. The task of 3D reconstruction is usually performed via a collection of color cameras and/or depth sensors. Such approaches thus require a synchronization protocol and a significant budget for device purchase. This report provides an overview of our system that employs only one depth camera together with two mirrors to perform the 3D reconstruction. The device cost is thus reduced and the problem of synchronization can also be avoided. Our approach is appropriate for providing intermediate real-time 3D models of dynamic objects with acceptable quality for practical applications.

1 Context

In order to obtain more useful features for describing an object, vision researchers attempted to perform the task of 3D reconstruction as a preprocessing step. Therefore many 3D reconstruction approaches have been proposed. The principle of 3D reconstruction is the combination of object-related information captured under different view points. There are various systems dealing with that idea such as employing a depth camera together with a turntable [6], shape-from-silhouette, using a collection of depth sensors [1] and/or stereo cameras [2]. These systems have their own limitations such as working with only static

*nguyetn@iro.umontreal.ca
†hhhung@dut.udn.vn
‡meunier@iro.umontreal.ca

objects, reconstructed models containing redundancy, requiring a synchronization protocol, or high prices. Our approach is thus proposed to overcome these problems.

Our system consists of a depth camera and two mirrors that are placed around the object[1]. Each depth map captured by the camera thus contains the object surfaces seen from 3 different view points. There are two common depth estimation schemes: Structured Light (SL) and Time-of-Flight (ToF). Our system using a SL depth camera has been presented in [5]. This report briefly describes the other camera type (ToF) where the reconstructed point cloud has more details compared with the work [5] but requires an additional processing to reduce depth distortion. The depth distortion occurs due to unwanted multiple reflections with mirrors. For further information on this work, the reader can refer to the original paper [4].

2 Algorithms

The reconstruction algorithm attempts to provide a point cloud obtained by combining (1) the object surface that is directly seen by the camera and (2) two reflections of the two other surfaces that are behind the two mirrors. A post-processing is then performed to remove *unreliable* points, i.e. 3D points having wrong depth measurement due to multiple reflections [3, 4]. Since this algorithm works directly on the pixels of each captured depth map, this can be integrated into real-time systems and is thus appropriate to work with dynamic objects. The details of this algorithm are presented in Algorithm 1, in which pts is the depth map provided by the camera that contains only object points, mir_1 and mir_2 are two mirror plane equations, C is the camera position, and reg is a predefined 3D region containing the object.

3 Experiments

In order to assess our approach, we performed the proposed 3D reconstruction algorithms on easy-to-evaluate geometric objects. We started with an experiment using only one mirror to demonstrate our hypothesis on mirror-related depth distortion as well as our solution (i.e. Algorithm 1) that reduces such distortions. A flat board of size $30cm \times 30cm$ was employed in this experiment. In the demonstration with two mirrors (our proposed system), we used a cylinder with radius of $15cm$ together with the flat board to assess our reconstruction accuracy. The Root Mean Squared Error (RMSE) calculated on the collection of reconstructed 3D points and its fitted shape was used to indicate the reconstruction error. Table 1 and 2 present such errors in our experiments under various settings (in term of geometric relation between an object and mirrors).

[1] A different number of mirrors can be used if needed.

Algorithm 1 : Reconstructing raw point cloud (appropriate for dynamic objects)

```
procedure GetCloud(pts, mir_1, mir_2, C, reg)
    cloud ← null
    P^(3D) ← reproject(pts)                              ▷ 2D to 3D reprojection
    for each point P in P^(3D) do
        if P inside reg then
            cloud ← push(P)
        else if P behind mir_1 then
            P_r ← reflect(P, mir_1)
            if P_r not inside reg then
                continue                                 ▷ check another point
            end if
            P_2 ← reflect(P_r, mir_2)
            if CP < CP_2 then
                cloud ← push(P_r)                        ▷ reliable point
            end if
        else if P behind mir_2 then
            P_r ← reflect(P, mir_2)
            if P_r not inside reg then
                continue                                 ▷ check another point
            end if
            P_1 ← reflect(P_r, mir_1)
            if CP < CP_1 then
                cloud ← push(P_r)                        ▷ reliable point
            end if
        end if
    end for
    return cloud                                         ▷ Return object point cloud
end procedure
```

Table 1: Reconstruction errors (in centimeter) when using 1 mirror. The term *angle* indicates the geometric angle (in degree) between the board and the mirror surface, *measured* and *enhanced* correspond to distance between ground truth points (determined directly by the camera SDK) and reconstructed points before and after performing our solution, respectively.

Angle	91.13°	89.22°	87.82°	86.26	84.70	83.46°	82.50°	81.50°	77.44°	73.74°
Measured	91.29	93.03	93.22	93.80	94.63	94.71	96.22	98.45	96.42	98.49
Enhanced	02.22	01.94	02.03	01.71	01.77	01.87	01.73	01.78	01.83	01.59

Table 2: Reconstruction errors (in centimeter) when using a system of 2 mirrors. The term *distance* indicates the average distance between the object and the two mirrors.

Board	Distance	60.70	67.06	76.12	82.09	86.74	91.78	97.27	101.50
	Naive reflection	0.209	0.242	0.250	0.243	0.216	0.212	0.202	0.195
	Algorithm 1	0.196	0.215	0.197	0.187	0.193	0.180	0.180	0.158
Cylinder	Distance	53.96	62.46	65.79	70.82	74.66	78.50	84.26	90.57
	Naive reflection	0.446	0.430	0.374	0.352	0.404	0.342	0.399	0.541
	Algorithm 1	0.359	0.384	0.285	0.324	0.371	0.333	0.387	0.499

According to Table 1, the depth measured on distorted points were significantly incorrect since the average distance between them and the ground truth was nearly 1 meter. By applying our Algorithm 1 that performs unreliable point removal, the measurement error was reduced about 53 times (95.9cm compared with 1.8cm).

Similarly, Table 2 shows that our procedure of unreliable point removal enhanced reconstructed point clouds since it decreased the reconstruction errors.

As indicated before, our 3D reconstruction approach is to provide intermediate 3D models with an acceptable quality in practical applications. An example of point cloud representing a 3D human body using the Algorithm 1 is presented in Fig. 1. This method is currently applied in our works on gait analysis.

Figure 1: Example of a point cloud obtained in our setup for an application of gait analysis.

If the researchers use our reported information in their works, please refer this document as

T.-N. Nguyen, H.-H. Huynh and J. Meunier, "Using ToF camera and two mirrors for 3D reconstruction of dynamic objects", Technical report 1380, DIRO, University of Montreal, April 2018.

References

[1] E. Auvinet, J. Meunier, and F. Multon. Multiple depth cameras calibration and body volume reconstruction for gait analysis. In *Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on*, pages 478–483, July 2012.

[2] G. K. M. Cheung, S. Baker, and T. Kanade. Visual hull alignment and refinement across time: a 3d reconstruction algorithm combining shape-from-silhouette with stereo. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–375–82 vol.2, June 2003.

[3] D. Freedman, Y. Smolin, E. Krupka, I. Leichter, and M. Schmidt. *SRA: Fast Removal of General Multipath for ToF Sensors*, pages 234–249. Springer International Publishing, Cham, 2014.

[4] T.-N. Nguyen, H.-H. Huynh, and J. Meunier. 3d reconstruction with time-of-flight depth camera and multiple mirrors. *IEEE Access*, 6:38106–38114, 2018.

[5] T.-N. Nguyen, H.-H. Huynh, and J. Meunier. Matching-based depth camera and mirrors for 3d reconstruction. In *Three-Dimensional Imaging, Visualization, and Display 2018, SPIE conference on*, volume 10666, pages 10666 – 10666 – 10, Orlando, FL, USA, April 2018. SPIE.

[6] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan. Scanning 3d full human bodies using kinects. *IEEE Transactions on Visualization and Computer Graphics*, 18(4):643–650, April 2012.