ABSTRACT

In many types of 3D cameras, The Time-of-Flight (TOF) cameras have the advantages of simplicity for use and lower price for general public. The TOF cameras can obtain depth maps at video speed. However, the TOF cameras suffer from low resolution and high random noise. In this paper, we propose methods to reduce the random noise in depth maps captured by the TOF cameras. For each point in the noisy TOF depth map, we substitute the depth value with the median depth value of its corresponding points in temporally consecutive depth maps captured by the TOF cameras. The proposed methods require only the depth data captured by the TOF cameras without any extra information, such as illumination, geometric shape, or complex parameters. Experiments results suggest that the proposed temporal denoising methods can effective reduce the noise in TOF depth maps for up to 44 percent.

Keywords Time-of-Flight camera; 3D data, Median; Random noise; Denoising;

1. INTRODUCTION

Recently, 3D data has more common usages for general public and are more widespread for different domains; e.g., 3D model reconstruction [2, 14]. 3D models are popularly used in many kinds of fields. 3D models of real-world objects are reconstructed with 3D data, such as the depth information or the color information, are collected by 3D scanners. Not only for 3D models, 3D data is also employed for applications such as collision prevention, position determination, real-time object detection, etc. Nowadays, people are concerned about read-time response. The speed of operation and the accuracy of the 3D information always influence the performance of 3D applications.

There are many selections of 3D scanners. Depending on different hardware and scanning technologies, each 3D scanner comes with its own advantages, shortcomings, and costs. Some scanners can produce high quality data, but they are quite expensive and often require expert knowledge for using. Users of those kinds of scanners are limited. We conjecture that if 3D cameras are easy to use and less expensive that will produce more applications and usages.

The TOF camera [2, 3, 8, 13] is one of the 3D scanners which are easy-to-use and less expensive. The type of the TOF camera is active, which means the camera emits an extra energy to the subject, and then using the Time of Flight principle measures the data of 3D space by the projection of the energy. The TOF camera can obtain the 3D depth map of a scene at video rate by measuring the traveling time of the light pulses between the camera and the objects. Furthermore, for the general public, the TOF camera is not difficult to operate and with a lower cost compares to the 3D laser scanners. Even though, the TOF camera is now more popular to industrial purposes than commercial usages. With the price dramatically decreases, the TOF cameras becoming off-the-shelf products, such as webcams or personal digital cameras, would not be a dream.

Although, there are many advantages of the TOF camera, it still suffers from several limitations such as random noise and lower-resolution depth maps. Some researches endeavor to improve the low-resolution problem of TOF camera [3, 14, 15, 16, 17, 19] by using one single depth map with adjacent depth pixels or by combining several low resolution noisy depth images of a static model. The proposed methods have some disadvantages, such as complex parameters need or long executive time. Other researches enhance the resolution and improve the distortion of 3D data by combining image data and depth data [6, 7, 11]. They still suffer from artifacts and use extra information, such as the intensity edges or the geometric shapes.

In this paper, considering as the basic of any processes, we suppose that the less noise depth maps have, the more accurate the information of depth maps is for TOF applications. Therefore, we focus on reducing the random noise in the depth maps of the TOF camera. We denoise the TOF depth maps only with the information provided in the TOF depth maps without any extra data from other sensors. Besides, the denoised method does not need to calibrate any complex parameters. Through the denoised methods we proposed, the TOF depth maps have more accurate depth values...
when comparing to the ground truth. The execution time of the proposed system is also fast and does not influence the advantage of using the TOF cameras.

In the remaining part of this manuscript, we elaborate on the TOF datasets, the denoising methods, and the experiment results.

2. TOF DEPTH MAP DENOISING METHODS

We first give a brief introduction of the TOF datasets employed for the experiments. After the descriptions, we elaborate each procedure of the proposed TOF depth map denoising method.

2.1. TOF Datasets

Y. Cui et al. [18] provide the TOF datasets we used in the experiments. A MESA Swissranger SR4000 TOF camera [8] captured the original depth maps from subjects in the datasets by its LEDs in the front of camera emitting light pulses and the lens of camera gathering the reflection of light to its CMOS imaging sensor for measuring the distances. The TOF camera rotated around one subject and captured 600 depth map frames, as shown in Figure 1. The angle between the first frame and the last frame is 120°.

![Figure 1: TOF Camera rotates around the rotation axis](image)

An example of the original TOF depth map transformed into a gray scale image is as shown in Figure 1.(a), where the depth values in the depth map are scaled to 0 to 255 gray values. The face in this image is severely corrupted by noisy pixels. Figure 1.(b) shows the 3D point cloud of the same TOF depth map in a 3D coordinate system. We observe that some points are very far from the surface of the 3D model and the outline of the model is indistinct. The TOF depth map suffering from these noising points would influence the accuracy of later processes. To reduce the noises in a TOF depth map, we modified the distorted depth values in a depth map which we treated as an X-Y plane with points and each point has its own depth value.

2.2. TOF Depth Map Denoising Processes

Median filter [1, 5] is one of the common smoothing methods for image denoising, which runs through the whole noise image pixel by pixel and replaces each pixel with the median value of the neighboring pixels. The number of the neighboring pixels involved is called “window”. The median filter has good performance of denoising and preserving edges with a fixed window size. Instead of using spatially neighboring pixels of each pixel in the original TOF depth map, we propose to apply the median filter with points which are corresponding to the point to be modified in temporally adjacent TOF depth maps.

To obtain a denoised TOF depth map, firstly we selected a frame Dt, where t is the number of the depth map in time. We assume that Dt’s preceding maps and succeeding maps were relevant to Dt. We then picked k consecutive depth maps, including the frame Dt, which k is the size of “window”. The frame Dk is the middle one of these k depth maps in time. The (k-1)/2 preceding maps and the (k-1)/2 succeeding maps of the Dt are aligned to middle map Dk and find the corresponding points from the depth maps for each point in Dk. The depth value of each point of each map is represented as Dk(x, y), where x is the x-coordinate and y is the y-coordinate of the point. Each point in the middle map has k corresponding depth values, including its own.

We compute the median of these k values which is the modified value of Dk(x, y). We refer to the proposed TOF depth map denoising method, which includes the multi-map TOF depth maps denoising process with the TOF depth maps alignment process, as the Temporal-Median.
The flowchart which describes the procedures of the Temporal-Median is shown in Figure 3. First, k consecutive TOF depth maps with the middle one $D_i$ are selected as input. Because of the rotation of camera, the same X-Y coordinate points in the k input TOF depth maps do not exactly indicate to the same position in the 3D scene represented by the TOF depth maps. By employing the TOF depth maps alignment process, each point in $D_i$ would find k aligned depth points with k adjusted depth values separately in k TOF depth maps. After alignment, denoising process compute the proper values for each point in $D_i$ with its own k aligning depth values and modify the depth value of each point in $D_i$. Finally, the Temporal-Median method outputs a denoised TOF depth map.

![Flowchart of TOF Depth Maps Alignment](image)

**Figure 3: Flowchart of Temporal-Median.**

### 2.3. TOF Depth Maps Alignment

In our TOF datasets, the TOF camera rotated around the object in the scene for 3D reconstruction purpose. This means that two points having the same X-Y coordinate in two TOF depth maps represents different position in the real 3D scene. Since the TOF camera did not move vertically when it rotated around the object, we only need to consider the X-Z coordinate offsets. Camera rotating around the rotation axis is equivalent to the object rotating around the axis in the converse wise. There are 600 TOF depth maps of frames in one dataset. The approximate angle of two frames is 2 degrees which is estimated with 120 degrees between the first frame and the last frame. Since that $D_i$ is the middle map of its k corresponding depth maps, the X-Z coordinate of each depth point in preceding maps is adjusted by turning around with the specific rotation axis in clockwise direction and which in succeeding maps is adjusted by turning around with the same axis in counterclockwise direction.

The proposed TOF depth maps alignment process has two stages. The first stage is to estimate the specific rotation axis with three X-Z coordinates of the particular feature point captured separately from three of depth maps. Firstly, three X-Z coordinates of the particular feature point in three depth maps of frame $A_1$, $B_1$, $C_1$, $(X_{a1}, Z_{a1})$, $(X_{b1}, Z_{b1})$, and $(X_{c1}, Z_{c1})$) are separately represented the X-Z coordinate of the particular feature point $a$, $b$, $c$ in depth maps of frame $A_1$, $B_1$, $C_1$, $(X_p, Z_p)$ is the X-Z coordinate of a specific point $p$ on the rotation axis $L$, and point $a$, $b$, and $c$ separately forms a straight line with this specific point $p$ which is on the rotation axis $L$. The angle of each straight line and the rotation axis $L$ is 90 degree. The distance between two points of the X-Z plane can be found using the distance formula. The distance between point $a$, $b$, and $c$ and point $p$ is represented as $l_a$, $l_b$, $l_c$, respectively. The distance from each point to the rotation axis $L$ is the same in any frame so that $l_a = l_b = l_c$. Using $l_a = l_b$ and $l_b = l_c$, we can derive Equation 1,

\[
\begin{align*}
(X_a - X_p)^2 + (Z_a - Z_p)^2 &= (X_b - X_p)^2 + (Z_b - Z_p)^2 \\
(X_b - X_p)^2 + (Z_b - Z_p)^2 &= (X_c - X_p)^2 + (Z_c - Z_p)^2
\end{align*}
\]

where $X_p$ and $Z_p$ are unknowns. According to Equation 1, a linear system is constituted as

\[
\begin{align*}
-2X_a + 2X_b - 2Z_a + 2Z_b &= X_p \\
-2X_b + 2X_c - 2Z_b + 2Z_c &= X_p \\
-2Z_a + 2Z_b - 2Z_b + 2Z_c &= Z_p
\end{align*}
\]

The linear system in Equation 2 is solved by linear algebra and thus the two unknowns, $X_p$ and $Z_p$, are acquired, which is the X-Z coordinate of the specific point $p$ on the rotation axis $L$. With the X-Z coordinate of point $p$ and the characteristic of the rotation axis $L$ which is perpendicular to X-Z plane, we can get the position of $L$ in 3D space.

The second stage of the TOF depth maps alignment is to find the k corresponding depth values of the k corresponding points of each point in $D_i$ with the specific rotation axis which is found from the first stage. Figure 4 shows the three steps of finding k corresponding depth values. To rotate adaptively, we transfer the rotation axis to the Y axis by translating the set of points of each depth map at first step. Second, as $D_i$ is the middle depth map which does not need to rotate, the preceding maps of $D_i$ turn around with the Y axis in clockwise direction and the succeeding maps of $D_i$ turn around with the Y axis in counterclockwise direction. After rotation, we find the k corresponding depth values of k corresponding points in k calibrated TOF depth maps for each point in $D_i$. 
Transfer the rotation axis to the Y axis

Calibrating the position of each point by rotating around the Y axis

Find the k corresponding depth values of each point in D_i

Figure 4: three steps in the second stage of the TOF depth maps alignment process

We then transfer the specific rotation axis L to the Y axis in each TOF depth maps, where p is a point on L and \((X_p, Y_p, Z_p)\) is the coordinate of point p. For all points in k TOF depth maps, referring their 3D coordinates to a 1D matrix and adding the 1D matrix \(M = [-X_p, 0, -Z_p]\) to each 3D coordinate is to move the specific rotation axis to Y axis. If the original 3D coordinate of one point \(c\) is \((X_c, Y_c, Z_c)\), after adding the matrix M, we have

\[
[X_c', Y_c', Z_c'] = [X_c - X_p, Y_c, Z_c - Z_p].
\] (3)

With all points in the k TOF depth maps adding matrix M, their rotation axis would be the Y axis.

Next, we calibrate the position of each point in the k TOF depth maps by rotation around the Y axis with its specific angle. The angle \(\theta_i\) of TOF depth map D_i for rotation is determined by the difference between i and t, which is represented by \(\Delta_i\), where i is the frame number of map D_i and t is the frame number of the middle map D_\(\text{m}\) of k TOF depth maps. For all i which is included in window k, the angle \(\theta_i\) is:

\[
\theta_i = 0.2 \times \Delta_i.
\] (4)

If D_\(\text{m}\) is one of preceding maps of D_i, it turns around the Y axis in clockwise direction using rotation matrix. If D_\(\text{m}\) is one of the succeeding maps of D_i, it turns around the Y axis in counterclockwise direction. After the rotation process, we have k calibrated depth maps.

In the final step, we capture k corresponding depth values of each point in D_i with k calibrated depth maps. For each point t in D_i, we find the corresponding point p in each calibrated depth map D_\(\text{m}\) by measuring the minimum distance of each pair of points. In equation (5), \(\text{cor}(t, p)\) is the corresponding point p in D_\(\text{m}\) of point t in D_i.

\[
\text{cor}(t, p) = \min_{p \in D_i}(\text{dist}(t, p))
\] (5)

\[
\text{dist}(t, p) = \sqrt{(X_t - X_p)^2 + (Y_t - Y_p)^2}
\] (6)

2.4. Denoising Process

The last stage of the Temporal-Median method is the denoising process for D_i by modifying each point in D_i according to its k corresponding depth values. In this paper, we employ the median filter with k corresponding depth values from the last stage to substitute the depth values in D_i. For each point in D_i, we sort its k corresponding depth values numerically. And, we replace the value of each point in D_i by the median value of its k corresponding value which are sorted, so that we can have a denoised TOF depth map. We also substitute different number into k, which is the window size of median filter, to have good performance of denoising. The denoised depth maps produced by the proposed denoising process, with either window size nine or window size twenty five, have less noise and the outlines of the model are obviously clearer than the original frame.

In addition to median filter, we also experiment with other smoothing methods to replace the values of points in D_i. We refer to this method as the Temporal-Mean method. For each point in D_i, instead of the median number of its k corresponding depth values, we use the mean value of k corresponding depth values to replace the value of point in D_i.

3. EXPERIMENTS

We experiment with several TOF datasets [18] and compare the denoised results with ground truth models, which are provided from the opening online laser scanning models [18] captured by a Minolta Vivid 3D scanner. Some examples of the ground truth models are as shown in Figure 5.
To evaluate the performance of the proposed Temporal-Media method and the Temporal-Mean method, the average difference of the depth values between the laser scanning data and noisy TOF depth map is employed. The original data captured by the TOF camera, compares with the average difference of depth values between the laser scanning data and denoised TOF depth maps produced by the proposed methods. In addition, we also try to deal with single TOF depth map by finding the respondent value of each point measured by Median filter or Smoothing method with an adjusted window size, which is the number of its nearest neighboring points in the same map. We also measure the average difference of depth values between the laser scanning data and the process data with the single TOF depth map.

The following results show the comparison of the average differences by different denoising procedures with the Head datasets. Figure 6 shows that denoising one example TOF depth map of Head dataset by Temporal-Median method with 9 window size and 25 window size. According to the figures of denoised results, the amount of points which are far from the model in Figure 6.(c) is less than in Figure 6.(b) and the outline of the models is clearer too.

We also try to denoise the original depth maps in Head datasets with single TOF depth map by computing the median value with 9 values and 25 values of nearest neighboring pixels. Figure 7.(a) is the denoised result, with window size 9, and Figure 7.(b) is denoised with window size 25. Comparing the denoised results of the Temporal-Median and the denoised results here, instead of clearer outline, with the bigger window size, the detail parts of the model in denoised results disappear and the outline of the model become blurred.
Figure 7: one denoised TOF depth map of Head dataset by single TOF depth map denoising method with different window sizes

Table 1 summarizes the denoised results with the Head dataset by different denoising methods and with two different window sizes. The average difference means that the average of depth value differences between the denoised depth map and the ground truth. And the improvement, which is compared with the original noisy TOF depth map and the denoised TOF depth map, indicate the deduction of the average differences in percentile. Table 1 show that denoising TOF depth maps with temporal consecutive depth maps result over 40% improvement. Denoising TOF depth map with spatially nearest neighboring points in one single depth map has only about 20% improvement. Table 1 also shows that when increasing the window size from 9 to 25, the Temporal-Median method produces additional 5% improvement. But with the spatial denoising methods, increasing the window size produces less than 1% improvement or even worse than with smaller window size.

Table 1: the average differences and the improvements of the denoised results with the Head dataset

<table>
<thead>
<tr>
<th>Denoising Method</th>
<th>9 window size</th>
<th>25 window size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average difference (cm)</td>
<td>Improvement</td>
</tr>
<tr>
<td>The original TOF Depth Map</td>
<td>0.83939</td>
<td>-</td>
</tr>
<tr>
<td>Temporal-Median</td>
<td>0.51301</td>
<td>38.883%</td>
</tr>
<tr>
<td>Temporal-Mean</td>
<td>0.49234</td>
<td>41.345%</td>
</tr>
<tr>
<td>Spatial-Median</td>
<td>0.66829</td>
<td>20.384%</td>
</tr>
<tr>
<td>Spatial-Mean</td>
<td>0.65556</td>
<td>21.899%</td>
</tr>
</tbody>
</table>

4. CONCLUSION

3D information has widespread usages in common applications, not only for distance detection, but also for 3D model reconstruction. 3D models are often used in many kinds of domains, such as industrial design or game industry. TOF camera is one kind of ideal 3D scanners which are user-friendly and not so expensive gradually. Besides, TOF camera can obtain the 3D depth map of a scene at video rate. However, there are still some drawbacks of TOF camera. One of them that we are concerned about is the high random noise depth maps produced by the TOF cameras.

In this paper, we propose TOF depth maps denoising methods using temporally consecutive depth maps captured by the TOF camera. Experiments demonstrate that the proposed temporal denoising methods not only produce more accurate depth maps, but also have the advantage in speed.

In future work, the boundaries of the models in TOF depth maps have some space for improvement. We will also try to incorporate the proposed denoising process into the 3D reconstruction framework to enhance the accuracy of 3D model reconstruction.

Acknowledgments

This work was partially supported by the National Science Council, Taiwan, under the Grants No. NSC101-2221-E-011-141, NSC100-2221-E-011-121, and NSC101-2221-E-211-011.
REFERENCES