REAL-TIME ACQUISITION AND SUPER-RESOLUTION TECHNIQUES ON 3D RECONSTRUCTION

Jong Wan Silva, Leonardo Gomes, Karl Apaza Agüero, Olga R. P. Bellon, Luciano Silva

IMAGO Research Group, Universidade Federal do Paraná (UFPR)
http://www.imago.ufpr.br

ABSTRACT

This work proposes to improve a traditional 3D reconstruction pipeline by combining it with two techniques: A real-time 3D modelling system to give a visual feedback in acquisition stage; and a super-resolution spatio-temporal filter to improve the depth data. We use noisy RGBD images from an inaccurate real-time depth sensor device in the whole process and later replace the color with data from a digital camera to generate realistic texture. Recently, several reconstruction approaches were presented acknowledging the potential of such real-time devices. However, those systems alone fail to retrieve small geometric characteristics from the object. The result of our experiments show a model with a considerable level of details, unexpected from low-quality RGBD images. We aim to use our pipeline to reconstruct objects with cultural value and add them to the virtual museum’s database for visualization purpose.

Index Terms— real-time, super-resolution, 3D reconstruction

1. INTRODUCTION

High resolution scanners were used in some of the finest works of digital preservation [1, 2]. Still, two of the main drawbacks are the slow capturing process and the lack of visual feedback on the acquisition stage, thus we depend on the reconstruction to confirm if an object was fully captured or if additional images will be required [3].

With the advent of new accessible and real-time depth devices, like Microsoft’s Kinect or Asus Xtion PRO, many new solutions for 3D reconstruction appeared in the literature [4, 5, 6, 7]. Although these techniques present fast and robust result systems they aren’t meant to capture small details in geometry and texture since these new devices present a low-quality depth information, but they’re still appealing considering the low-cost and real-time nature.

In many cases a more accurate solution like a 3D laser scanner isn’t available or even possible due to restrictions in acquisition scenario. For this reason we extend Vrubel et al. [3] 3D reconstruction pipeline, a complete 3D modelling system focused on digital preservation of cultural heritage, to work with real-time depth sensor devices and overcome the scanners limitation.

The proposed pipeline can be understood as follow. Initially we acquire real-time RGBD images and high resolution color images of an object using a depth sensor and a digital camera respectively. To guide the acquisition, we use an implementation of KinectFusion [4] to create a real-time updating 3D model used only for visual feedback. Then, we apply a super-resolution method [8] to improve the devices raw data.

Next, all the processed RGBD images are merged into a final 3D model with the adapted version of Vrubel et al. [3]. And finally, the texture is replaced with the digital camera images, like in Andrade et al. [9].

We demonstrate the effectiveness of our approach comparing it with the real-time state-of-the-art 3D reconstruction, KinectFusion.

2. RELATED WORK

Real-time systems. In these past years, interesting works have been proposed for the sake of real-time 3D modeling. Henry et al. [5] presented an algorithm that combines color feature in the Iterative Closest Point (ICP) pose estimation, resulting in a more robust alignment. The model is rendered with surfels, suited for dynamic geometry. However, the registration process performs an average of 500ms. Neumann et al. [6] implemented the color ICP on the graphic processing unit (GPU) with the novel Random Ball Cover (RBC) [10], a nearest neighbour search that exploits the parallel architecture of modern graphics cards, achieving ICP runtimes of 20ms.

Currently, the KinectFusion [4] is the real-time 3D reconstruction system with the best result in the literature. With a GPU-supported implementation, it combines a fast ICP with the volumetric Truncated Signed Distance Function (TSDF) representation to create accurate dense models. The TSDF range can be improved with Whelan’s et al. [7] approach, creating maps of extended scale environments.

Thanks to UFPR, CAPES and CNPq for funding.
Super-resolution. A simple and efficient noise reduction technique is the bilateral filter [11], an edge-preserving filter that uses spatial information to improve a single image. Since low-cost devices suffer from random noises, other works like Cui et al. [12] take advantage from the temporal information to merge several aligned depth images and obtain a single depth with improved resolution.

To avoid relying on the smoothness prior alone for the super-resolution, some other works also exploit the combination of depth with color images to correct and improve the depth information. Kopf et al. [13] proposed the Joint Bilateral Upsampling (JBU) to upsample a low-resolution image using additional information available in another high-resolution image. Yang et al. [4] creates a cost volume with the high-resolution image and then apply a bilateral filter.

In Snively et al. [15] approach, several steps are used to define a spatially and temporally smooth directional field over the combined depth and color data. The results are accurate, but the combined data is somewhat difficult to capture, because of their expensive equipment.

An improvement of Snively et al. work is proposed by Richardt et al. [8], a super-resolution approach designed for real-time performance that handles the challenging noise characteristics of recent depth cameras, producing high-quality depth maps at interactive frame rates.

3D Reconstruction. The 3D reconstruction is a very active research area, and presenting entire pipelines there are several works like Bernadini et al. [16], Ikeuchi et al. [2], Curless et al. [17] and Levoy et al. [1]. Vrubel et al. [3, 9], combine the qualities of the previous mentioned approaches and is also successfully employed on cultural heritage preservation projects that demands high geometry and texture accuracy.

Cui et al. [12] also present interesting results with real-time devices, however, their work focus more on time-of-flight scanners and don’t take advantage of a real-time feedback for the acquisition.

3. PROPOSED APPROACH

Our approach combines the efficiency of a real-time capturing system with the high quality of an offline reconstruction process. To achieve that, we combine three distinct steps:

1. A real-time system to reconstruct a low-quality 3D model that will guide the user during the acquisition.

2. A spatio-temporal super-resolution technique to improve the quality of our sensor’s depth images.

3. An offline reconstruction process using the super-resolved data to create a high-quality 3D model with improved texture from a high resolution digital camera.

Figure 1 presents our complete pipeline, the steps are detailed in the following sections.

3.1. Real-time acquisition

A remarkable feature of real-time systems is the ability to align, reconstruct and render the model at high frame rates, overcoming common capturing problems. This is important for the first step of our pipeline, since we want a fast and complete scan of the object.

In our work, we use the KinectFusion’s open source implementation, released by the Point Cloud Library (PCL) project [18] with support to a variety of sensing device.

It starts with a bilateral filter [11] to provide better results for the surface vertex and normal map pyramid during the surface measurement step (a.1). This information is applied in the ICP for fast pose estimation (a.2). We take advantage of this process and save all aligned information for the super-resolution stage. Next, we incrementally fuse each consecutive frame into a single 3D reconstruction using TSDF (a.3). This volumetric representation allow us to perform a per pixel raycast (a.4) to render the real-time 3D model.
3.2. Super-resolution

We apply Richardt et al."s [8] super-resolution approach in our work. Their method first re-projects the depth data to the color image, since this last one is a more reliable information (b.1). Then, a fast fill-in procedure handles half-occluded regions (b.2). Finally, a spatio-temporal filter that also incorporates information from previous frames denoises and super-resolves the depth images (b.3), improving its resolution. The color images are used to increase the spatial-resolution of depth images by exploiting the coincidence of color and depth edges. Since we are dealing with a large amount of data (30 to 60 frames per second) of processed images, we select the middle image of each set of ten consecutive images for the future reconstruction.

Although this method is able to improve the depth information with high-resolution color image, optical flow tends to be unreliable due to motion blur in areas of fast motion.

3.3. 3D Reconstruction

After the previous super-resolving techniques, an integration process is needed to fuse all the processed views into a complete high-quality digital model. For this step, we decided to use an offline reconstruction process (the Vrubel et al.'s [3] reconstruction) for the following reasons:

- The final reconstruction doesn't need to be in real-time.
- The chosen system is a set of state-of-the-art methods successfully employed in 3D reconstruction of cultural heritage preservation projects [19].

The system first pre-aligns the processed depth data with a SURF-based technique [20], followed by a two-phase ICP algorithm: a point-to-plane convergence followed by a point-to-point minimization, this last offering maximum precision (c.1). It's important to mention that those data were aligned before in the real-time step. But to avoid possible failures from the previous fast ICP, we align once more.

For the mesh integration we use the improved version of the Volumetric Range Image Processing (VRIP) and Consensus Surfaces, the IMAGO Volumetric Integration Algorithm (IVIA) [21], to combine all views (c.2). Holes caused by incomplete data acquisition are filled with a volumetric diffusion algorithm (c.3), and the mesh is generated with the Marching Cubes (c.4).

For the texture, we calibrate the high resolution color images to depth images [20] (c.5). In the texture parametrization, we use an atlas approach which cuts the mesh in regions with low curvature (c.6). These regions are parametrized and packed in a texture atlas together with the surface properties (c.7) to generate the final texture (c.8).

Optionally, the final model can be simplified (c.9) for improving the rendering performance, storage costs and still maintain the high visual quality [22].

4. EXPERIMENTAL RESULTS

In our experiments, the acquisition has been performed using Microsoft Kinect with resolution of $640 \times 480$ for depth frames and $1280 \times 960$ for color, and after the super-resolution step both receive a $1066 \times 833$ resolution. Approximately 1000 frames were acquired and processed, and 100 of them used to generated 3D models with our approach.

The reconstruction processing time for each object takes around 10 minutes under average PC settings.

4.1. Geometry discussion

Given the high precision presented by the 3D laser scanners and the lack of an appropriate evaluation method, we established the models reconstructed with Konica Minolta VIVID 910 laser scanner combined with IVIA reconstruction as a ground truth for comparison purpose. In figure 2, we present the results of three different approaches from two objects: (a)(b) The laser scanner model; (c)(d) the state-of-the-art KinectFusion; (e)(f) our approach.

Our approach preserved some geometric details evident in figure (e) marked in green, which are similar to the scanner model (a). The equivalent image from KinectFusion (c) over-smoothed the geometry which result in loss of details. The same effect can be seen in the Indonesian board sculpture images (b)(d)(f).

A limitation of our approach is caused by the ambient light which can affect the super-resolution results. This can interfere the geometry of each super-resolved depth image, and the differences in these images also affect the registration process giving a noisy aspect in some regions marked in a red square in the figures (e)(f).

We also evaluate the Root Mean Square (RMS) distance error of KinectFusion and our approach against the scanner model. The results can be seen in the table 1, and shows that our approach slightly outperforms the real-time reconstruction. Even with these positive outcome, we understand that RMS alone is not a reliable evaluation method.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Penguin (b)(d)</th>
<th>Board (a)(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KinectFusion</td>
<td>1.995482</td>
<td>1.259788</td>
</tr>
<tr>
<td>Our approach</td>
<td><strong>1.866729</strong></td>
<td><strong>1.225292</strong></td>
</tr>
</tbody>
</table>

Table 1. RMS distance error comparison between the Kinect generated models against the scanner.

4.2. Texture discussion

The proposed pipeline uses color images of $4368 \times 2912$ resolution provided by a professional digital camera Canon EOS 5D, but for evaluation purpose we create another model with
Fig. 2. 3D model of a penguin, generated with (a) Minolta scanner, (c) KinectFusion and (e) our approach; and an Indonesian board sculpture generated with (b) Minolta scanner, (d) KinectFusion and (f) our approach. The green squares represent the details preserved in our approach (e)(f) which is similar to the scanner result (a)(b). The KinectFusion preserved the object’s shape, but over-smoothed the model (c)(d). The limitation of our approach is the registration problem that affect model’s shape, marked with red squares.

Fig. 3. Indonesian board sculpture texture: The first image is a photo of the object; second image is the a rendered model with the texture obtained from kinect; and the third image using texture from digital camera.

5. CONCLUSION

In this paper, we present some improvements for a 3D reconstruction pipeline, extending it to work with real-time depth sensor devices. To assist the RGBD images acquisition stage we apply a real-time reconstruction method to generate a low-resolution model only to guide the acquisition process. After that, we apply super-resolution methods in the raw data and merge our RGBD images and high resolution color images into a single mesh in a post process. Our experiments shows an impressive level of geometric details but also present some deformation that can affect the registration. Combined with a high quality texture image, we managed to create models adequate for visualization in virtual museums. In the future, we plan to solve the deformation problem using non-rigid registration approaches and improve overall quality for art preservation applications.

6. REFERENCES


