J. Vis. Commun. Image R. 21 (2010) 427–441

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

A quality controllable multi-view object reconstruction method for 3D imaging systems

Wen-Chao Chen [a], Hong-Long Chou [b], Zen Chen [a,*]

[a] Department of Computer Science, National Chiao Tung University, 1001 University Road, Hsinchu 300, Taiwan
[b] Altek Corporation, Science-Based Industrial Park, Hsinchu, Taiwan

ARTICLE INFO

Article history:
Received 2 July 2009
Accepted 22 February 2010
Available online 21 April 2010

Keywords:
3D imaging system
Modeling from silhouettes
Octree model
XOR projection error
System performance
Dynamic modeling
Progressive transmission
Multi-camera system

ABSTRACT

This paper addresses a novel multi-view visual hull mesh reconstruction for 3D imaging with a system quality control capability. There are numerous 3D imaging methods including multi-view stereo algorithms and various visual hull/octree reconstruction methods known as modeling from silhouettes. The octree based reconstruction methods are conceptually simple to implement, while encountering a conflict between model accuracy and memory size. Since the tree depth is discrete, the system performance measures (in terms of accuracy, memory size, and computation time) are generally varying rapidly with the pre-specified tree depth. This jumping system performance is not suitable for practical applications; a desirable 3D reconstruction method must have a finer control over the system performance. The proposed method aims at the visual quality control along with better management of memory size and computation time. Furthermore, dynamic object modeling is made possible by the new method. Also, progressive transmission of the reconstructed model from coarse to fine is provided. The reconstruction accuracy of the 3D model acquired is measured by the exclusive OR (XOR) projection error between the pairs of binary images: the reconstructed silhouettes and the true silhouettes in the multiple views. Interesting properties of the new method and experimental comparisons with other existing methods are reported. The performance comparisons are made under either a comparable silhouette inconsistency or a similar triangle number of the mesh model. It is shown that under either condition the new method requires less memory size and less computation time.

1. Introduction

The three-dimensional imaging technology is an emerging research topic for capturing, processing and displaying the true 3D information of a scene object. The next generation of 3D imaging systems allows spectators to view from any desired viewpoint, just like in the real world. Two common multi-view solutions to the next generation of 3D imaging systems are: (1) novel view synthesis and (2) multi-view 3D modeling [1–5]. The main challenges of the emerging 3D imaging systems include 3D model reconstruction, video transmission rate, and 3D free viewpoint rendering, etc.

The stereoscopic display based on view synthesis is a practical approach for 3D imaging nowadays. The depth-image-based rendering (DIBR) methods for free viewpoint TV (FTV) were introduced in [3–5]. Virtual images were rendered from a small number of views with 3D warping. Many researchers also utilized video interpolation technique to synthesize novel views without building the 3D shape [6–8]. Although these kinds of methods can provide high quality images by interpolation, the viewing angle is limited by the initial camera positions.

In contrast to view interpolation, multi-view 3D modeling approaches construct the 3D geometry of the scene and, therefore, are more suitable for applications in the holographic 3D display or 3DTV systems, which require to render views from all directions, not just the in-between views. The 3D mesh with texture mapping is the most common approach for producing the photorealistic objects. In recent years, there has been an increasing amount of literature on image-based 3D modeling. Multi-view stereo algorithms were proposed to reconstruct the high quality 3D model from images captured at multiple viewpoints [9–14]. Seitz et al. [15] used high quality multi-view datasets as the benchmark to evaluate the performances of different reconstruction algorithms based on accuracy and completeness of the reconstructed objects. Multi-view stereo algorithms are generally based on photometric consistency measurement which is a time consuming procedure.

On the other hand, visual hull is an alternative approach to 3D modeling using multi-camera systems [16–19]. Franco and Boyer [21] addressed the exact polyhedral visual hull reconstruction by cutting and joining the visual rays passing through silhouettes. Despite the complexity in joining the visual ray segments, the

* Corresponding author. Fax: +886 3 573 1875.
E-mail addresses: chaody.cs94g@nctu.edu.tw (W.-C. Chen), hlchou@altek.com.tw (H.-L. Chou), zchen@cs.nctu.edu.tw (Z. Chen).

1047-3203/$ - see front matter © 2010 Published by Elsevier Inc.
doi:10.1016/j.jvcir.2010.03.004

resultant visual hull is highly accurate in terms of silhouette consistency. However, the constructed polyhedral visual hull tends to produce ill-formed triangles (irregular and narrow planar surface strips) [35]. Liang and Wong presented a simple and efficient way to compute the exact visual hull vertices by the exact intersection computation which replaces the interpolation value used in the conventional marching cubes (MC) approach [35]. This exact visual hull from marching cubes was reported to significantly improve the quality of the reconstructed visual hull, comparable to that of the polyhedral approaches and requiring less computational time. Nevertheless, the method demands a pre-specified subdivision level number for the octree reconstruction. If the level number is too small, the octants generated intend to have a larger volume, resulting in lower accuracy, while if the level number is too large, requiring a tremendous computation time and memory space. On the other hand, the parallel execution of the voxel-based visual hull reconstruction is feasible. In [25] Ladikos et al. proposed a method making use of CUDA to perform the reconstruction by using kernels which perform the projection of every voxel and gather the occupancy information. They also showed a real-time 3D reconstruction system which used the GPU-based visual hull computation algorithm. There is a website where the information on CUDA is available [26]. Starck et al. [22] matched surface features between views to modify the visual hull and applied a global optimization approach to do a dense reconstruction. Vlasic et al. [23] and de Aguiar et al. [24] acquired a template body mesh which could be generated by a laser scanner and then tracked the motion sequences by deforming the non-rigid body mesh shape to match faithfully with a video stream of silhouettes.

This paper proposes a novel octree reconstruction method. The conventional approach often faces with a conflict between model accuracy and memory size. Since the tree depth or the subdivision level is discrete, these system performance measures (accuracy, memory size, and computation time) are varying rapidly. To remedy this drawback, we modify the conventional method to attain much finer control over the system performance. In the new method the visual quality of the octree reconstructions is controlled by a specified exclusive OR (XOR) projection error upper bound; the resultant XOR projection error reflects the inconsistency between the binary projected silhouette of the reconstructed object and the true object silhouette in each image. The introduction of new types of octants in the new reconstruction method indicates a mixture of protrusions and indents on the reconstructed object surface which is no longer a bounding volume of the true object. Both useful properties and computer simulations of the system performance are presented. Furthermore, dynamic object modeling is made possible based on the new method. Also, progressive transmission of the reconstructed model with an increasing degree of model accuracy is provided. Since the proposed method is fast and has the dynamic and progressive nature, together with a controllable system performance measure, the method is suitable to be incorporated into 3D imaging systems. Comparisons between the proposed reconstruction method and other existing methods are made to illustrate the merits of the new method.

The main features of our work include

(1) The proposed method does not enforce a maximum tree depth for the octant generation process; instead an XOR projection error upper bound parameter is imposed. This parameter value selection depends roughly on the intended level of detail. Therefore, the user has some sort of control over the visual reconstruction quality.
(2) Under a comparable XOR projection error (i.e., silhouette inconsistency) constraint the proposed octree reconstruction method is faster than the conventional octree reconstruction method by a factor from 10 to 40. The total processing time of our method including the conversion of the octant-base octree to a surface mesh representation is faster than the conventional octree method with the standard marching cubes method for generating the final surface mesh model (the "Conv + MC" method) and the method of generating an exact visual hull from marching cubes based on the standard octree model (the "Conv + ExMC" method or simply the ExMC method) by a factor from 2 to 3. Furthermore, the reconstructed visual hull contains a triangle number for the surface representation which is less than that of the existing methods.
(3) The proposed method can generally achieve better accuracy, while spending less computation time through the introduction of new octant types to the conventional octree ones.
(4) Due to the application of exact marching cubes the reconstructed visual hull mesh is relatively smooth, so it does not need further surface smoothing for texture mapping.

The rest of the paper is organized as follows. Section 2 briefly describes the framework of a proposed multi-view capturing and processing integrated system for generating the 3D texture mapped objects. Section 3 presents useful properties of the new reconstruction method with regard to its system performance. In Section 4 the experimental setup is described. Synthetic and real objects of different geometry complexity are used in the visual hull reconstruction. The system performance measures of the proposed reconstruction method are given. Comparisons of our reconstruction method with other existing methods are provided. Conclusions and further research directions are given in Section 5.

2. The proposed system

To fulfill the requirements of FTV systems or 3DTV applications, a multi-view capturing and processing system generates the photorealistic appearance of 3D dynamic objects from multi-view images based on some image-based reconstruction algorithm. Fig. 1 shows the overall system organization. Yemez and Schmitt also presented a similar efficient approach to progressively transmit the octree structure and then perform triangulation on the fly [38]. In our system, the preprocessing module first comprises camera calibration, background estimation, and color normalization, etc. After capturing a synchronized video sequence, the system applies background subtraction for 3D model reconstruction, and finally the multi-view texture mapping is derived for photorealistic display of the 3D object. Only octree structure information and multi-view images are transmit from a server to clients. Here, the focus is placed on the extension of the conventional octree reconstruction and the application of exact marching cubes to convert the octant-based volumetric representation to surface mesh representation of the object under reconstruction. Also, dynamic modeling and progressive transmission based on the new method are to be presented. Texture mapping and Z buffer will be implemented with the commercial graphic cards.

2.1. A multi-view capturing system

We implement a multi-view capturing system consisting of eight IEEE 1394b synchronized color cameras which are connected to eight PCs. The PCs are synchronized with an IEEE 1394a daisy chain. The system configuration is shown in Fig. 2. Only a single computer command is needed to trigger the capturing process through the communication interface mechanism. The cameras capture a video sequence at a rate of 30 fps with a frame resolution of 1024×768. We also set up a blue screen-like studio of dimension $3\ m \times 6\ m \times 2\ m$ and all the cameras are mounted on the