



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.

© 2010 Published by Elsevier Inc.

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), hichou@altek.com.tw (H.-L. Chou), zchen@cs.nctu.edu.tw (Z. Chen).

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-based 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 $3\text{ m} \times 6\text{ m} \times 2\text{ m}$ and all the cameras are mounted on the

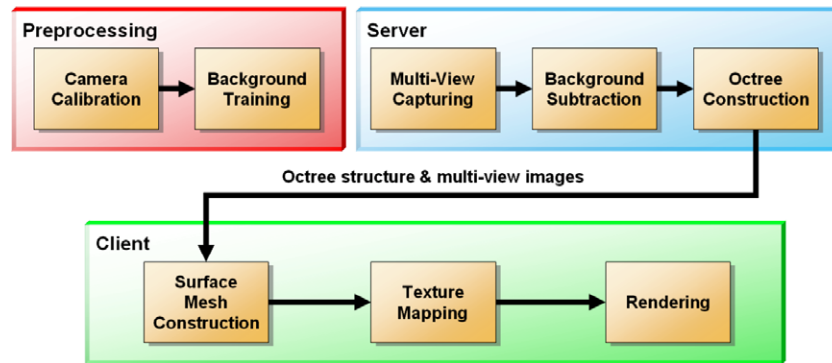


Fig. 1. System block diagram.

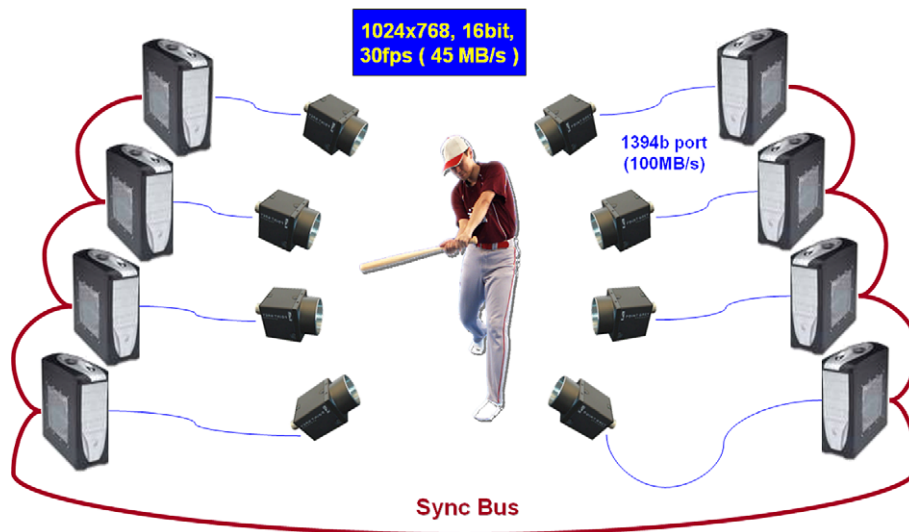


Fig. 2. A multi-view capturing system.

ceiling to acquire the images of the objects from different viewing positions. Fig. 3 shows our studio configuration.

A color checker chart is used to estimate the true color information for color normalization and a custom-made calibration board is used for acquiring the projection matrix for each camera based on the Bouguet's method [28]. Although objects are captured in a blue screen environment, foreground extraction is still not easy due to shadows casted by the objects. This paper combines the methods proposed in [30–31] with the manually selected threshold to extract foreground objects for the 3D reconstruction.

2.2. A novel octree reconstruction method with a projection error control

In the conventional octree reconstruction a coarse-to-fine octant subdivision process is recursively applied to a bounding root cell of an object under reconstruction. During the subdivision process the black nodes indicate the octants lying inside the object, while the white nodes indicate the octants lying outside the object and the gray nodes indicating the octant lying partially outside and partially inside the object. The octant subdivision process is



Fig. 3. Multi-view images captured.

repeated until no gray octants remain. A typical octree representation of an object model reconstruction is shown in Fig. 4. In the conventional octree reconstruction method, the type of an octant is classified by checking the intersection relation between the projected octant image and the real object silhouette in each view. In order to speed up the intersection test, an octant is approximated by its bounding sphere and the intersection test is executed using precompiled signed distance maps derived from the object silhouettes [27]. The signed Chebyshev distance map for each silhouette view can be generated with a graphic card to reduce computation time using the method proposed by Hoff et al. [32]. The distance value at the circle center c of the bounding circle of a projected octant image is denoted by $DistMap(c)$. The radius r of the bounding circle of the projected octant image is calculated and then compared with $DistMap(c)$ to determine the intersection relationship. A positive $DistMap(c)$ distance value indicates the circle center is inside the object silhouette, and a negative distance value for a circle center being outside the silhouette, and a zero value for a circle center being on the silhouette boundary. The spatial relationship checking is carried out on the view-by-view basis. Table 1 gives the intersection relationship between the projected octant image and its associated silhouette.

To measure the model accuracy of the reconstructed visual hull O relative to the true object T , we use their binary silhouettes S_v^O and S_v^T across all views ($v = 1, \dots, M$) to define the following average exclusive OR error:

$$XOR\ error = \frac{1}{M} \sum_{v=1}^M S_v^O \otimes S_v^T \quad (1)$$

where $S_v^O \otimes S_v^T$ stands for the total number of the XOR errors between pixels of the two binary silhouettes in each view v .

Another way to measure the reconstruction model accuracy is to use the percentage of the error pixels of the reconstructed object given by the following ratio:

$$Err(O, T) = \sum_{v=1}^M S_v^O \otimes S_v^T / \sum_{v=1}^M Area(S_v^T) \quad (2)$$

where $Area(S_v^T)$ is the area (or the total pixel number) of the region of the true silhouette. For a good object reconstruction this ratio should be much smaller than one.

Note that the error caused by the approximation with a bounding circle becomes negligible as the subdivision level increases to a sufficiently large final level. Also, the computation of the above XOR projection error uses the actual hexagonal projected images of the reconstructed object octants rather than using the bounding circles of the octants.

In the new octree reconstruction method a pre-specified XOR octant projection error bound is denoted by P . There are five types of octants at each resolution level l in the new method: $B_{l,N}$ (black), $W_{l,N}$ (white), $GB_{l,N}$ (gray-black), $GW_{l,N}$ (gray-white), and $GG_{l,N}$ (gray-gray). The black and white octants are defined in the conventional way and the three new types of gray octants are defined below:

A conventional gray octant $G_{l,c}$ at level l is redefined as a gray-white ($GW_{l,N}$) octant in the new method if its bounding circle center c has a negative distance value, $DistMap(c) < 0$ and the black extent of the octant satisfies the constraint: $0 < r + DistMap(c) \leq P$ in any particular view $v \in [1, M]$.

A conventional gray octant $G_{l,c}$ at level l is redefined as a gray-black ($GB_{l,N}$) octant in the new generalized method if its bounding circle center c has a non-negative distance value, $DistMap(c) \geq 0$ and the white extent of the octant satisfies the constraint: $0 < r - DistMap(c) \leq P$ in all views $v \in [1, M]$. If a gray octant $G_{l,c}$ cannot be redefined as one of the above two, it is redefined a gray-gray ($GG_{l,N}$) octant. That is, white and black extents of a $GG_{l,N}$ octant both exceed the octant projection error bound P .

A relatively large portion of a gray-white octant remains outside the silhouette whereas no significant portion of a gray-gray

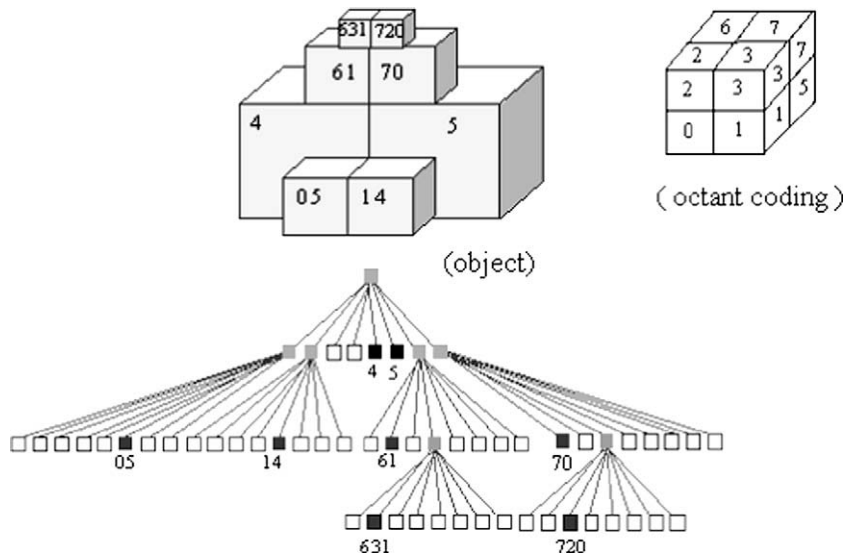


Fig. 4. A typical octree representation of an object model.

Table 1

Intersection relationship between the projected octant image and its associated silhouette.

$DistMap(c) \geq 0$	$r > DistMap(c)$ in any view $r \leq DistMap(c)$ in all views	The projected octant image intersects the silhouette (indicating a gray octant) The projected octant image is within the silhouette (indicating a black octant)
$DistMap(c) < 0$	$r > -DistMap(c)$ in any view $r \leq -DistMap(c)$ in any view	The projected octant image intersects the silhouette (indicating a gray octant) The projected octant image is outside the silhouette (indicating a white octant)

octant is either outside or inside the silhouette. With the new definition of the gray octant types the octant subdivision scheme of the new generalized reconstruction method is given as follows.

In the new generalized octree method at each level $l = 0, 1, \dots, L_{N,P}$ only the $GG_{l,N}$ octants need to be subdivided into eight child octants. The octant subdivision process is performed from level to level until there are no new $GG_{l,N}$ octants at the next (finer) level. The final level, $L_{N,P}$, of the new method with a specified XOR octant projection error bound P is defined as the level at which the set of gray–gray octants generated by the new method is empty.

In a previous paper of ours [20] we extended the conventional black octant to include the gray–black (GB) octant so that the octree generated is still a bounding volume of the real object. It has been shown that if a proper XOR projection error bound is selected, the extended method spends less memory space and processing time, while achieving a comparable visual quality of the conventional method with a properly fixed subdivision levels.

In this paper we generalize the conventional octree method to include both the gray–black (GB) octants and gray–white (GW) octants. The presence of GB octants indicates the protrusion of the reconstructed object model and the presence of GW octants indicates indent or shrinkage of the reconstructed object model. In other words, the generalized model produces protrusions and indents on the object surface and the new generalized model is no longer a bounding volume of the true object. As shall be seen, the generalized octree method reduces memory space and processing time of the system compared to the aforementioned methods under the comparable silhouette inconsistency condition.

To our best knowledge, Erol et al. are the only group concerned with the conditional gray octant split problem [36]. They proposed a condition for determining whether a boundary (i.e., gray) octant is split further or not. The condition depends on the length of the diagonal of bounding rectangle of the voxel projection and the minimum local feature size (LFS) value inside the bounding rectangle. These two values change with the voxel under consideration, so the boundary voxel's split decision is not simple. In our method the boundary voxel's split depends on a fixed projection error bound P , so it is much easy to verify and the error bound parameter can be set according to the level of detail wanted. From this point of view the octree reconstruction of our method is generally faster than the adaptively sampling method.

2.3. A dynamic refinement of the reconstructed octree model with respect to a decreasing error bound parameter P

After the octree model is reconstructed by the new reconstruction method with a given octant projection error bound P , the resultant octree model can be dynamically updated without the need of restarting from scratch when the parameter P changes to a smaller value. Namely, for the new smaller parameter P only those gray–black and gray–white octants associated with an octant's XOR projection error greater than the new P value are needed to split. In this way the octree model can be dynamically refined. The refinement process can be recorded as the split of a parent gray–black or gray–white into eight child octants each with a smaller octant's XOR error. The forward and backward pointers can be used to log the octant split history.

2.4. A progressive transmission of the reconstructed octree model with a best-first traversal scheme

In the octree reconstruction process the XOR projection error of each generated octant is evaluated. The information on the projection error can be used not only for the dynamic reconstruction (as described above) but also for progressive transmission. For transmission of the reconstructed object model, a priority queue, called

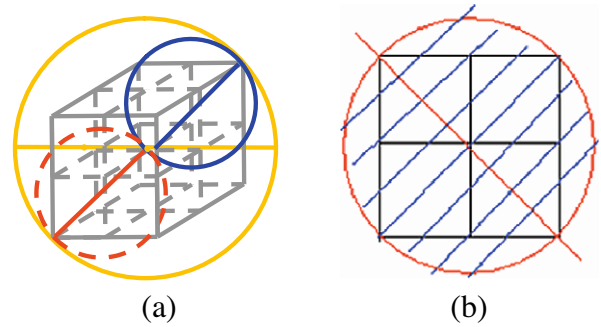


Fig. 5. (a) The bounding spheres for the parent and child octants at two consecutive levels. (b) The top view of the parent and eight child octant images shown with a bounding circle diameter and a number of object cutting edges.

priority_queue, is used to store the sorted list of the unprocessed octants according to the decreasing order of the octant's XOR projection error. Each time when an octant subdivision is requested, the first octant in the *priority_queue*, whose projection error value is currently the largest, is fetched for subdivision and marked as a processed octant with the associated projection error bound P . After the subdivision the eight sub-octants are inserted into the *priority_queue* based on their XOR projection error values. Also, the octant split history of the progressive model reconstruction is recorded by the forward and backward pointers.

During the model transmission only the gray–black octants and gray–white octants are required to be transmitted for rendering. The progressive transmission is terminated when the first octant in the *priority_queue* has an error value no greater than the specified error bound P .

2.5. Surface mesh representation and texture mapping

In the progressive transmission mode the reconstructed octree model and one or more multi-view images are transmitted to the client. Then a marching cubes technique is used on the client to obtain the exact intersection of each projected octant edge with the silhouette in each image, if there is an intersection. Among the corresponding intersection points found from all images the one closest to its silhouette in all views is chosen as the final exact point on the visual hull surface. In this way the octant-based volumetric representation is converted into a surface representation in a mesh form. The accuracy of our reconstructed visual hull mesh is



Fig. 6. The hardware setup of the multi-view capturing system with a turntable connected to a PC.

basically similar to that of the “Conv + ExMC” method except that the former starts with a generalized octant model and the latter a conventional octant model.

Cautions are exercised to prevent the reconstructed mesh from cracking. In a reconstructed 3D model by our method, the neighboring octants may be in different levels and in different sizes. In other words, our octant model is not the same as the conventional one which contains the gray octants at the final subdivision level only. Cracks occur when the marching cubes method is applied to octants at different resolution levels. In order to prevent cracks, currently we split non-terminal octants at lower levels into descendant octants all the way to the final level prior to the application of the marching cubes technique, although it may not be necessary. In this way no cracks will occur. A Poisson surface reconstruction algorithm (PSR) developed by Kazhdan et al. [33] can be employed to produce a fairly smooth visual hull mesh. However, PSR is a time consuming technique and is not suitable for the real-time 3DTV applications.

For a single-user client only one rendered view is required for the given user viewing angle. Choosing the one or two images captured around the user viewing angle would be appropriate for texture mapping at the client end. Given the viewing angle, the UV texture mapping coordinates of each vertex on a visible triangular mesh can be computed using the known camera projection matrix. Texture mapping can be done in less than 30 ms by many off-the-shelf graphic cards.

3. Properties of the new generalized octree reconstruction method

In the following the notation $GG_{l,c}$ stands for a single gray-gray octant generated at level l in the conventional method or it stands

for the entire set of all gray-gray octants generated at level l , depending on the context. The corresponding notation in the new method is denoted as $GG_{l,N}$. The notations for other octants are similarly defined.

For the object reconstruction one has to specify the dimensions of the root octant (a cube) based on the input image resolution. One can compute the location of any octant inside the root octant given the dimension of the root octant and use the known camera projection matrix to compute the size of the octant's projection image. The bounding circles of the octants generated at the same level change slightly in size due to their different locations inside the root octant.

From the empirical experience, one often finds a typical octant subdivision outcome as follows. Starting from the root octant at level 0, eight or a slightly fewer number of gray sub-octants are generated at level 1. The new gray octants generated are recursively subdivided. When the level is sufficiently large (e.g., level > 4) any gray octant will be generally subdivided into 4 gray, 2 black and 2 white child octants. This pattern is called the 4–2–2 subdivision pattern hereafter.

Definition 1. The (object) octree memory space of the conventional method is determined by the space required to store the total number of black octants at all levels, $\sum_{l=1}^{L_c} |B_{l,c}|$, and the total number of gray octants at the final level L_G , $|GG_{L_G,c}|$. The object reconstruction time of the conventional method is defined as the total number of octants generated by the conventional method at all levels. (Namely the total octant number determines the actual octree reconstruction time.)

It is difficult to derive the precise object memory size and the object octree reconstruction time of the old and new reconstruction methods, but it is desirable to have a rule of thumb for



Fig. 7. Three real objects used in the experiments: (a) a face, (b) a dinosaur, and (c) a figure sculpture.

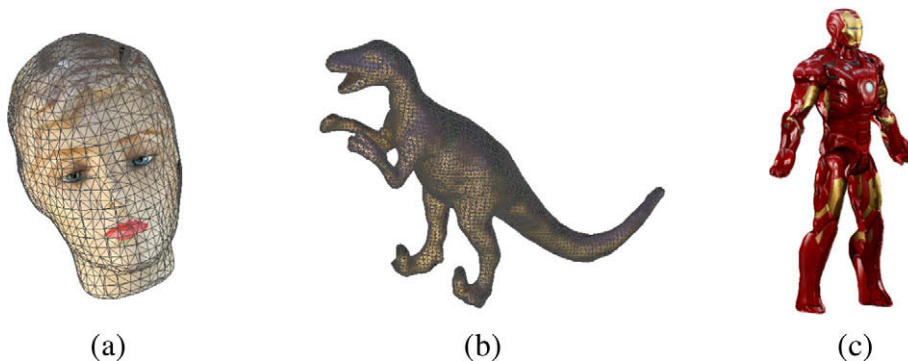


Fig. 8. New views generated from the reconstructed models.

Table 2

The statistics of octants generated from the real dinosaur dataset by two different reconstruction methods with their specified system parameters.

Octant generation level	Conventional method $L_C = 7$, XOR = 127,757					New method $P = 42$, XOR = 103,790				
	B	GB	GG	GW	W	B	GB	GG	GW	W
0	0	0	1	0	0	0	0	1	0	0
1	0	0	4	0	4	0	0	4	0	4
2	0	0	14	0	18	0	0	14	0	18
3	0	0	59	0	53	0	0	39	20	53
4	0	0	202	0	270	0	0	130	71	111
5	21	0	800	0	795	21	86	205	445	283
6	407	0	3107	0	2886	45	1138	0	396	61
7	3266	0	12,392	0	9198					
8										
$L_C = 8$, XOR = 66,228						$P = 18$, XOR = 45,918				
0	0	0	1	0	0	0	0	1	0	0
1	0	0	4	0	4	0	0	4	0	4
2	0	0	14	0	18	0	0	14	0	18
3	0	0	59	0	53	0	0	49	10	53
4	0	0	202	0	270	0	0	169	33	190
5	21	0	800	0	795	21	27	540	226	538
6	407	0	3107	0	2886	217	496	1131	1324	1152
7	3266	0	12,392	0	9198	433	4664	0	3344	607
8	18,934	0	48,086	0	32,116					
9										
$L_C = 9$, XOR = 33,852						$P = 12$, XOR = 32,961				
0	0	0	1	0	0	0	0	1	0	0
1	0	0	4	0	4	0	0	4	0	4
2	0	0	14	0	18	0	0	14	0	18
3	0	0	59	0	53	0	0	52	7	53
4	0	0	202	0	270	0	0	185	17	214
5	21	0	800	0	795	21	15	637	148	659
6	407	0	3107	0	2886	295	307	1801	959	1734
7	3266	0	12,392	0	9198	1178	3529	1818	5496	2387
8	18,934	0	48,086	0	32,116	489	11,405	0	2448	202
9	88,349	0	182,434	0	113,905					
10										
$L_C = 10$, XOR = 20,821						$P = 7$, XOR = 20,405				
0	0	0	1	0	0	0	0	1	0	0
1	0	0	4	0	4	0	0	4	0	4
2	0	0	14	0	18	0	0	14	0	18
3	0	0	59	0	53	0	0	56	3	53
4	0	0	202	0	270	0	0	193	9	246
5	21	0	801	0	794	21	8	704	89	722
6	410	0	3109	0	2889	347	153	2361	593	2178
7	3260	0	12,422	0	9190	2092	1931	6396	3758	4711
8	18,935	0	48,180	0	32,261	5733	16,991	2450	19,995	5999
9	88,425	0	182,990	0	114,025	1355	15,934	0	2156	155
10	397,159	0	600,509	0	466,252					

Table 3

The diameter range of the bounding circles of the projected octant images at all levels.

Range (unit: pixel)	Level									
	1	2	3	4	5	6	7	8	9	10
$D_{\min, h} D_{\max, l}$	[1590, 3176]	[698, 1398]	[350, 676]	[172, 336]	[90, 166]	[46, 80]	[22, 42]	[12, 22]	[6, 8]	[2, 8]

Table 4

The tabulation of the XOR_{C, L_C} errors, the average $(XOR_{C, L_C} + XOR_{C, L_C+1})/2$, memory space, reconstruction time of the conventional method with level $L_C = 6, \dots, 9$, respectively. The last row gives the error bound value P of the new method with XOR model error comparable to the conventional one at the respective level.

L_C	6	7	8	9	10
XOR_{C, L_C}	233,831	127,757	66,228	33,852	20,821
$(XOR_{C, L_C} + XOR_{C, L_C+1})/2$	180,794	96,993	50,040	27,337	–
M_{C, L_C} (in log scale)	3.55	4.21	4.85	5.47	6.04
T_{C, L_C} (in log scale)	3.94	4.53	5.12	5.71	6.30
P	–	83	40	18	9

calculating these figures for performance evaluation. A basic assumption made here is the 4–2–2 subdivision pattern in the conventional method, as explained below.

When the subdivision level is sufficiently large, any gray–gray octant will have a small volume and the real object surface intersects the gray–gray octant in a random fashion; half of the octant lies inside and half lies outside the real object. Also, half of the child octants will be white or black and half of the child octants are gray–gray, as illustrated in Fig. 5(b). As a consequence, the numbers of gray–gray, white and black child octants have a ratio of 4:2:2. And the area of the new XOR error due to the gray–gray

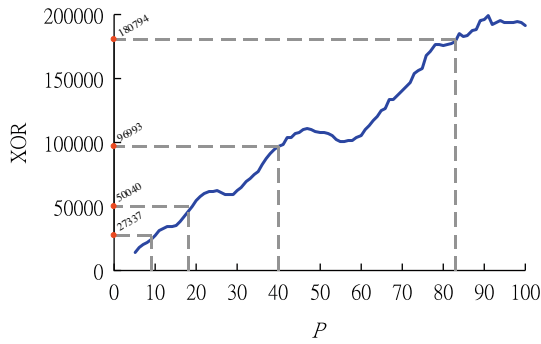


Fig. 9. The (dense) plot of the XOR projection error vs. the error bound parameter P . The horizontal dash lines indicate the average value $(XOR_{L_c} + XOR_{L_{c+1}})/2$ and the vertical dash lines indicate the corresponding error bound values P of the new method with an XOR model error comparable to that of the conventional method. The corresponding error bound values of P as indicated on the horizontal axis are given in Table 4.

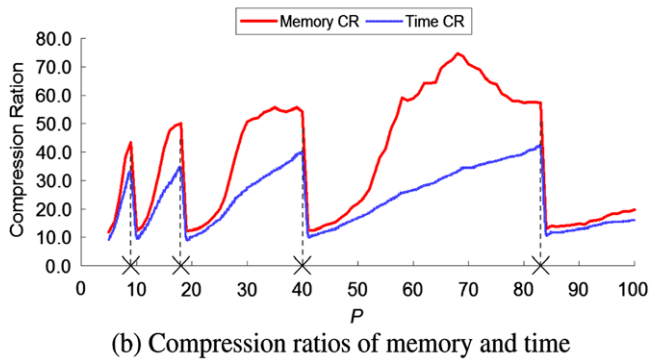
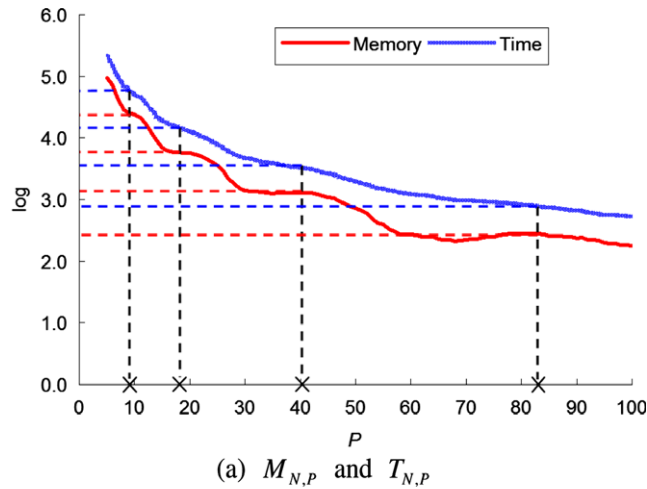


Fig. 10. (a) Plots of the memory size and reconstruction time (in the logarithm scale) vs. parameter P of the new method. (b) Plots of memory and reconstruction time compression ratios between the two reconstruction methods under the comparable XOR model accuracy.

sub-octant images is roughly reduced to about half of that of the parent gray-gray octant after the subdivision level is raised by one.

Therefore, the following property, **Property 1**, has been established:

Property 1. When the subdivision level becomes sufficiently large, a gray-gray parent octant is split into child gray-gray, white and black child octants in a ratio of 4:2:2

Property 2. At a sufficiently large final subdivision level of the conventional method the octree memory space and reconstruction time both increase approximately by four times as the level increases by one.

Property 3. At a sufficiently large final subdivision level of the conventional method the XOR projection error reduces roughly by half as the level increases by one.

The proofs of **Properties 2, 3** and the following **Property 4** are provided in **Appendix A**.

In order to compare the system performance of the reconstruction methods, a notion of reconstruction model accuracy comparability between the different methods is defined below:

Definition 2. Let XOR_{L_c} be the projection error of the octree model (see Eq. (1)) obtained by the conventional method with a final level L_c . If the new method with a given XOR projection error bound P has an XOR projection error lying in the range from $(XOR_{L_{c-1}} + XOR_{L_c})/2$ to $(XOR_{L_c} + XOR_{L_{c+1}})/2$, then these two methods are said to have the comparable reconstruction model accuracy (or silhouette inconsistency).

Definition 3. The object memory space of the new method is determined by the space to store the sets of black octants and gray-black octants at all levels generated in the method. The object reconstruction time of the new method is determined by the total number of all five kinds of octants generated at all levels in the method.

Property 4. The XOR model accuracy of the new method is nearly a non-decreasing function of the error bound parameter P except occasionally at a few ranges of P . On the contrary, the XOR model accuracy of the conventional method is a non-decreasing function of the subdivision level.

Property 5. Under the comparable XOR model accuracy condition the ratio of the object memory size of the conventional method to that of the new method, $M_{C,L_c}/M_{N,P}$, is found empirically to be in the interval of [12, 80]. Similarly, the ratio of the object reconstruction time of the conventional method to that of the new method, $T_{C,L_c}/T_{N,P}$, is found empirically to be in the interval [10, 40].

From above the new method has a better memory and time complexity than the conventional method under the comparable XOR model accuracy. This is mainly due to the introduction of the new octant types of gray-white and gray-black octants.

Table 5

Comparison between the XOR model errors of the three reconstruction methods under a similar triangle number.

	David (19 images)			Dinosaur (36 images)			Dancer (20 images)		
	Triangles	XOR error	$Err(O, T)$	Triangles	XOR error	$Err(O, T)$	Triangles	XOR error	$Err(O, T)$
Conv + MC	27,368	21,561	6.525%	62,616	147,671	7.139%	10,198	90,665	6.632%
Conv + ExMC	27,368	21,244	6.481%	62,616	120,586	5.830%	10,198	75,594	5.530%
Ours	26,602	15,786	4.888%	69,242	87,891	4.249%	10,897	73,897	5.406%

Table 6

Comparison between the numbers of triangles of the three reconstruction methods under a comparable accuracy.

	David			Dinosaur			Dancer		
	$Err(O, T)$	XOR error	Triangles	$Err(O, T)$	XOR error	Triangles	$Err(O, T)$	XOR error	Triangles
Conv + MC	8.007%	26,787	6,658	7.139%	147,671	62,616	5.572%	76,178	42,200
Conv + ExMC	7.367%	24,100	6,658	6.677%	138,103	15,192	5.530%	75,594	10,198
Ours	7.180%	23,189	6,350	6.089%	125,946	16,734	5.406%	73,897	10,897

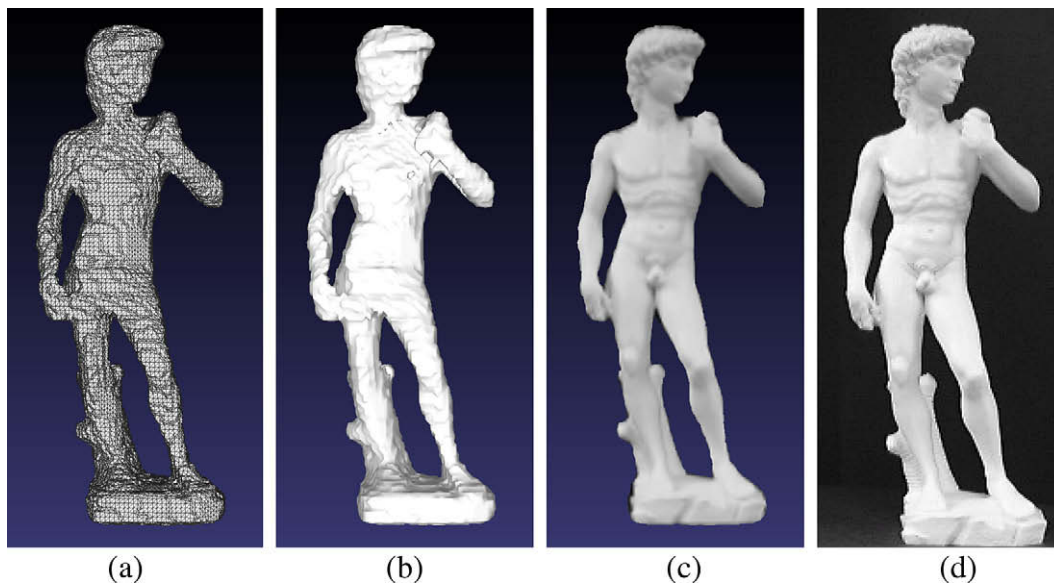
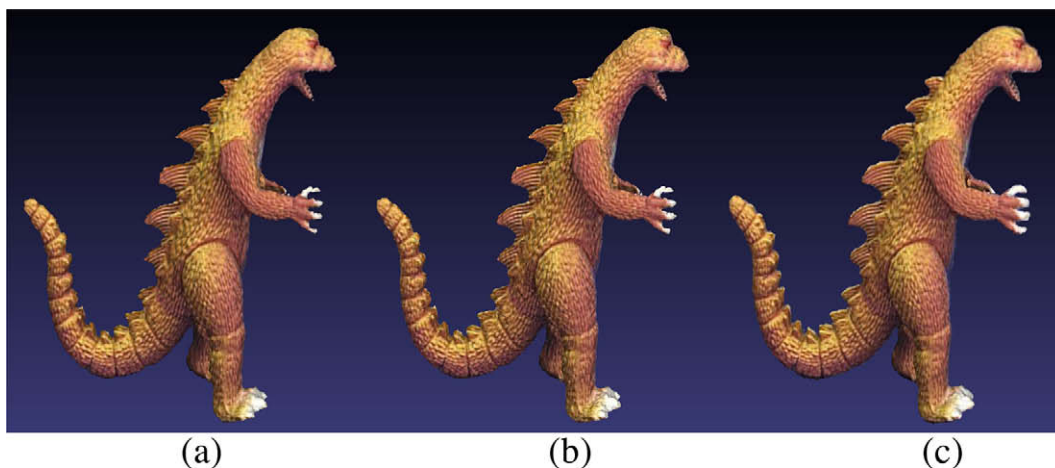
Table 7

Execution time (in milliseconds) comparison of the three reconstruction methods under a similar triangle number.

	David	Dinosaur	Dancer
Triangles	27k	62k	40k
Conv + MC	1360	4078	1765
Conv + ExMC	1453	4204	1728
Ours	750	1390	624

* Running on Intel Quad Core 2.83GHz Processor with 3GB Ram

However, there is a price to pay for the performance improvement. That is, when the XOR projection bound P is greater than the half size of the most elongated part (i.e., the finest detail) in the object model, and then it is possible that the reconstructed object model given by the new method may have a missing part. The missing part can be avoided by shifting the root octant or using a smaller value P in the new octree reconstruction method. In practice, the value P can be decided according to the projected size of the finer part in the object model.

**Fig. 11.** Reconstructed model representations for the David sequence. (a) Mesh representation, (b) 3D shaded representation, (c) texture mapping representation, (d) one of the original images.**Fig. 12.** The rendering results with texture mapping of the reconstructed dinosaur image sequence. (a) Conv + MC, (b) Conv + ExMC, and (c) our method. The results are found under a similar number of triangles.

4. Experimental results

To evaluate the proposed reconstruction method, we design several experiments on a variety of synthetic and real objects of different geometric complexity. The first experiment deals with real objects and the second experiment reconstructs 3D objects from synthesized images in different image resolutions. The third experiment shows the progressive reconstruction results of a real human frozen by the multi-view capturing system. We also demonstrate an augmented reality (AR) application using the reconstructed 3D models.

Fig. 6 depicts our hardware system including a turntable connected to a PC. The camera captures a sequence of images (10–36) of the real object resting on a rotating turntable under the control of a PC. The whole reconstruction program is written in VC++ under the Windows environment. Some typical real objects used are shown in Fig. 7 and the new views generated from the reconstructed 3D models are given in Fig. 8.

4.1. Real object reconstructions

To analyze the reconstruction results, 10 images of a real dinosaur resting on a turntable are taken in the experiment. Each image has a resolution of 2700×1800 pixels. The main intermediate results of the conventional method and the new method are collected below to show their performance differences.

Table 2 is the tabulation of the numbers of all types of octants generated by the conventional method with a specified subdivision level ($L_c = 7, 8, 9$, and 10) and by the new reconstruction methods specified with a comparable XOR octant projection error bound values ($P = 42, 18, 12$, and 7). The other fields in the table include “octant generation level” indicating the level when the octants are generated, and the sizes of the five generated octant sets: $|B|$, $|GB|$, $|GG|$, $|GW|$, and $|W|$. From this table one can see the subdivision patterns of gray–gray octants into various sub-octants during the whole subdivision process in these two different reconstruction methods. One can check the important properties of these

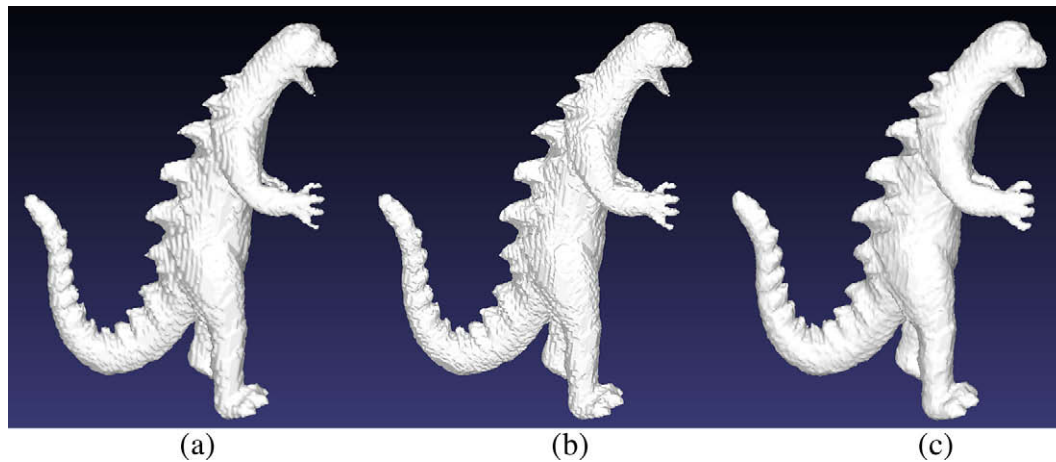


Fig. 13. The rendering results with shading of the reconstructed dinosaur image sequence. (a) Conv + MC, (b) Conv + ExMC, and (c) our method.

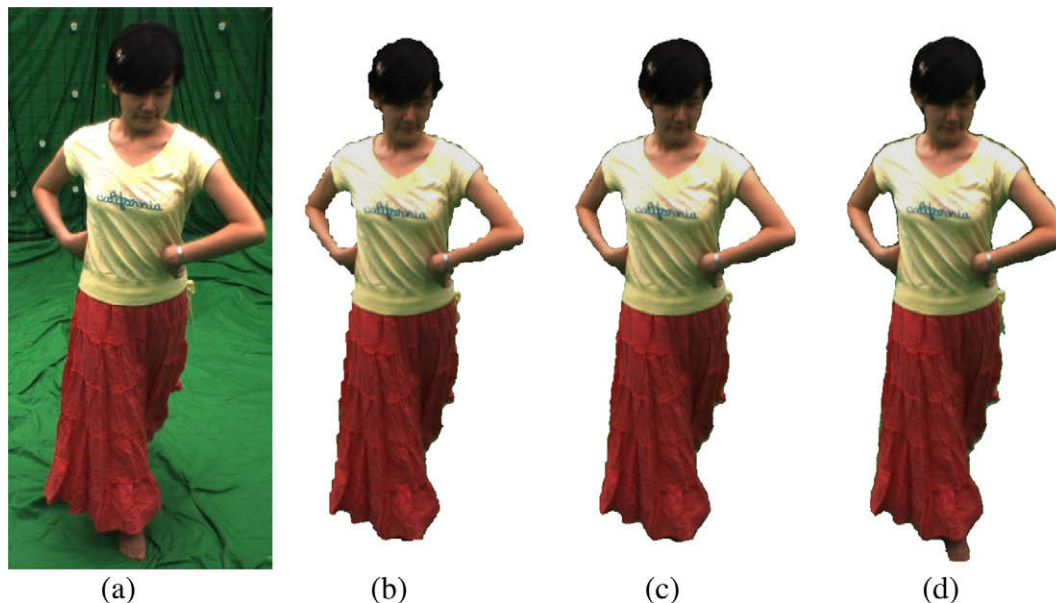


Fig. 14. The rendering results with texture mapping of the reconstructed dancer image sequence. (a) One of the original images (b) Conv + MC, (c) Conv + ExMC, and (d) our method. The results are under a similar number of triangles.

two methods described in the previous section including (i) the relative sizes of the various sets of octants of different methods at each level, (ii) the final level numbers of the new methods relative to that of the conventional method, and (iii) the XOR errors at all levels in these methods.

The root octant used in this experiment has a dimension of 50 cm^3 . The root octant is placed at the center of the turntable where the rotation axis passes through. As mentioned previously, any octant at level $0, 1, 2, \dots, L_C$ is approximated by a correspond-

ing bounding sphere. The projection of this sphere onto the image plane of each camera can be found using the camera projection matrix. The size of the projected circle varies with the sphere location in the space. The ranges of the minimum and maximum diameter $[D_{\min, l}, D_{\max, l}]$ of the projected circle can be obtained in advance. They are listed in Table 3.

These minimum and maximum values indicate the minimum and maximum projection error of the final object model obtained at the level l by the conventional reconstruction method. They will

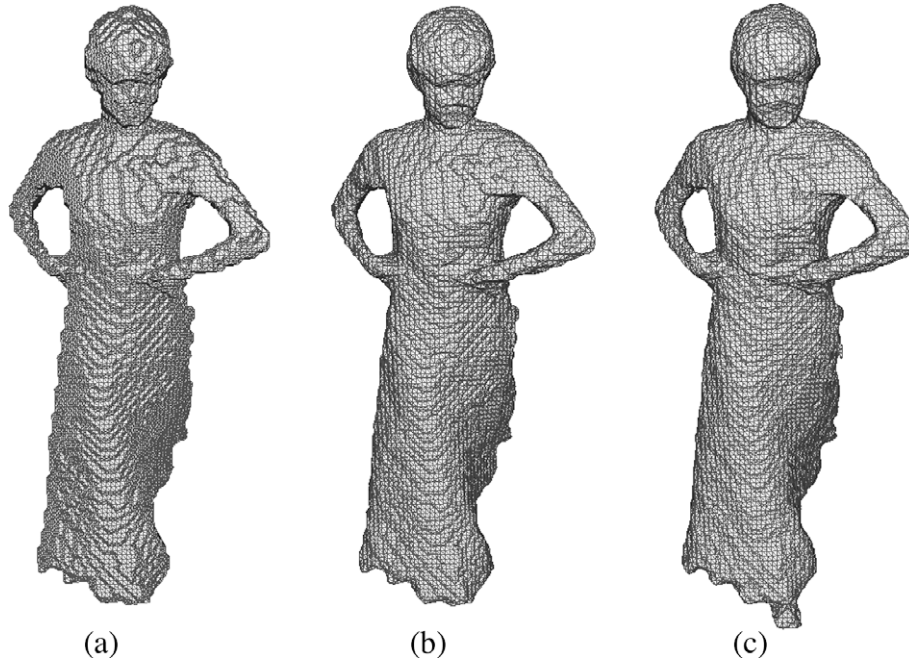


Fig. 15. The mesh rendering results of the reconstructed dancer image sequence. (a) Conv + MC, (b) Conv + ExMC, and (c) our method.

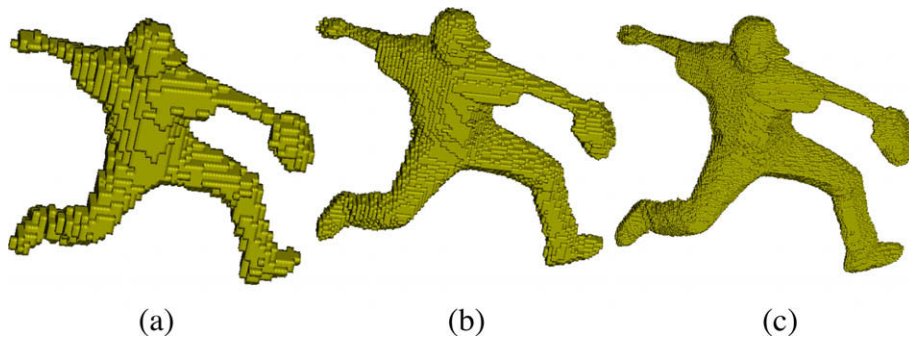


Fig. 16. Progressive reconstruction results of the octree model creation at the transmission instants corresponding to an XOR octant projection error bound P value equal to (a) $P = 13$, (b) $P = 7$, and (c) $P = 4$.

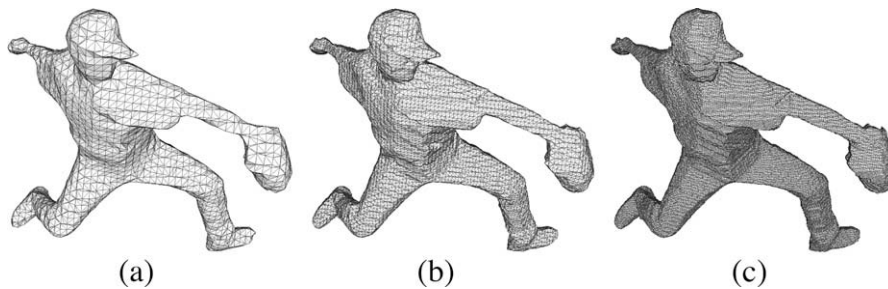


Fig. 17. Progressive reconstruction results of the visual hull mesh after applying the exact marching cubes technique.

be used to predict the feasible XOR octant projection error bounds P used in the new reconstruction method.

The XOR model accuracy of the two reconstruction methods are given in Table 4 and Fig. 9, respectively. The former is a discrete function of the final subdivision levels and the latter is a dense plot with respect to the error bound parameter P . In Table 4, the average values of the XOR model errors at any two consecutive levels are computed, as shown by the diamond marks along the vertical axis in Fig. 9. From these average values the selection of the corresponding XOR error bound values of P can be found, as indicated on the horizontal axis of Fig. 9. These correspondence data will be used below to compute the memory compression ratio and the time compression ratio between the conventional and new reconstruction methods for their performance evaluations. Recall that if the new method with an error bound P has an XOR model error

lying in the range from $(XOR_{L_c-1} + XOR_{L_c})/2$ to $(XOR_{L_c} + XOR_{L_c+1})/2$, then the new method and the conventional method are said to have the comparable reconstruction accuracy.

The logarithm values of memory space and reconstruction time for the conventional method at level $L_c = 6, 7, \dots, 10$ are given in Table 4. In Fig. 10(a) the logarithm values of the memory space (in red) and the reconstruction time (in blue) of the new method are listed for the range of P values; the dash vertical lines indicate those P values given in Table 4. Based on the data shown in Fig. 10(a) one can compute the memory compression ratio and the time compression ratio between the conventional and new methods, as shown in Fig. 10(b). In this experiment the memory compression ratio falls in the range of $[12, 80]$ and the time compression ratio falls in the range of $[10, 40]$. We also conducted the octree reconstruction using the images at higher image

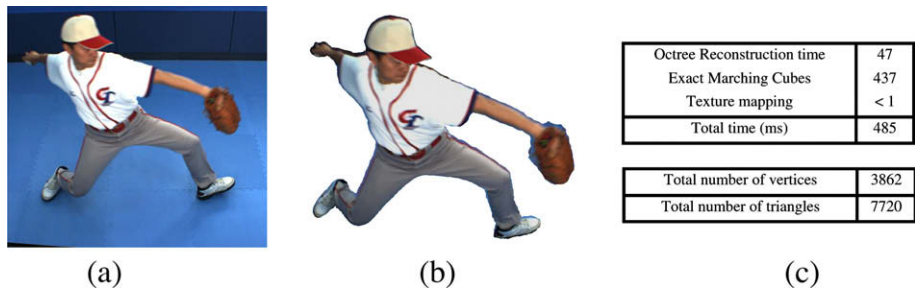


Fig. 18. (a) One of the original images; (b) the final textured mapped visual hull mesh in Fig. 17(a) after applying exact marching cubes technique as well as texture mapping; (c) the processing time of the system components and the total numbers of vertices and triangular faces, respectively.



Fig. 19. Free viewpoint visualization by making use of ARToolKit.



Fig. 20. Assignment of different actor models according to specific markers.



Fig. 21. Replay of the changing 3D models according to the recording time line.

resolution and observe that the ranges of memory and time compression ratios almost do not change with the image resolution.

In the next experiment we acquire from the public website three image sequences: David (19 images), dinosaur (34 images), and dancer (20 images) [37]. We implement three methods for the reconstruction of the subjects in these image sequences. They are the “Conv + MC” method [29], the “Conv + ExMC” method or, simply, ExMC [35], and our generalized octree reconstruction method with the exact marching cubes intersections. Tables 5–7 give the comparisons of the XOR model error, the number of triangles, and the execution time of the three methods under an associated comparability condition, respectively. From these tables one can see that the performance of our method ranks first, Conv + ExMC second, and Conv + MC last. The merit gap widens as the level of detail of the subject increases.

Figs. 11–15 show the reconstruction results of the three methods for the three image sequences in various model representations. The dinosaur models in Figs. 12(c) and 13(c) are locally different from the others. As indicated in Table 5 our reconstructed dinosaur result has a smaller XOR error value than those of the two other methods; therefore, our method achieves a better reconstruction quality. Generally speaking, our methods allows gray-black and gray-white octants so that our reconstructed visual hull approximates the true object from both inside and outside of the object surface, while other methods generally start with a bounding volume of the true object with more non-terminal octants prior to the marching cubes method. We can use different control parameter values of P to generate thicker or thinner reconstruction results.

4.2. Progressive reconstruction from real human images

The new reconstruction method combined with progressive transmission mechanism mentioned previously is used to model a real human in the multi-view capturing system. Fig. 16(a)–(c) show the progressive transmission of the reconstruction results using the best-first scheme and the projection errors of the generated octants are ranked in the decreasing order. Fig. 17 shows progressive reconstruction results. Fig. 18 shows the progressive reconstruction results with texture mapping. Our programs run on an Intel Quad Core PC with 3 GB RAM, but without any code optimization.

4.3. Applications to augmented reality

The system can use the augmented reality software, ARToolKit [34], to visualize the reconstructed 3D models in a more interesting way. The interactive AR application implemented on the host PC detects all possible patterns in the video frame and displays the correct motion sequences of baseball characters in accordance with the 2D markers captured by the camera. Thus, the user can hold and rotate the marker cards to enjoy free viewpoint visualization, as shown in Fig. 19.

Fig. 20. shows the designed interactive baseball game in which different reconstructed baseball characters are associated with dif-

ferent marker cards. Each character has his own motion such as the pitching or the batting. All the motions are pre-recorded and can be replayed according to their recording time line, as shown in Fig. 21.

5. Conclusion

This paper presents a multi-view capturing and processing system with a new generalized octree reconstruction method to generate an object visual hull. The conventional gray octants are refined into gray-gray, gray-black, and gray-white octants in the new method. The octree model produced by the new method approaches the real object from both the outer and inner object boundaries. The 3D model reconstructed provides good object model and photorealistic rendering effect. With the progressive transmission, users can get a quick preview of 3D object under reconstruction and obtain a refined 3D object as time goes on.

For interactive 3DTV applications, our experimental results show that it is possible to reconstruct the 3D object models in real time with a small number of PCs running in parallel. For the multi-view images transmission between the server and a client, only one or two images are needed to transmit for texture mapping at the client end.

Experimental results show that under the comparable XOR projection error constraint using our generalized reconstruction method can reduce the memory space required by the conventional method by a factor of 12–80 and the octree reconstruction time required by the conventional method by a factor of 10–40. Also, the experimental reconstruction results obtained from three image sequences available in the public websites indicate that our method can speed up the processing time by a factor of 2–3, when compared with the “Conv + MC” method and the “Conv + ExMC” method under a comparable silhouette inconsistency constraint.

Currently, our system is undergoing an optimization of its code to accelerate the processing speed to meet more stringent application need. Also, multi-view stereo registration is studied to do the dense reconstruction using the visual hull acquired as an initial object model.

Acknowledgments

We would like to express our gratitude to reviewers for their constructive comments which lead to the improvement of the paper. This work was supported by the Industrial Technology Research Institute of Taiwan (97C078, 98C052) and National Science Council of Taiwan (NSC97-2221-E-009-145).

Appendix A. Proofs of Properties 2, 3, and 4

Property 2. At a sufficiently large final subdivision level of the conventional method the octree memory space and reconstruction time both increase approximately by four times as the level increases by one.

- [27] R. Szeliski, Rapid octree reconstruction from image sequences, *Computer, Vision Graphics and Image Processing: Image Understanding* 58 (1) (1993) 23–32.
- [28] J.Y. Bouguet, Camera calibration toolbox for Matlab. Available from: <http://www.vision.caltech.edu/bouguet/calib_doc/>.
- [29] W.E. Lorensen, H.E. Cline, Marching cubes – a high resolution 3D surface reconstruction algorithm, in: *Proceedings of the SIGGRAPH*, 1987, pp. 163–169.
- [30] R. Zhang, S. Zhang, S. Yu, Moving objects detection method based on brightness distortion and chromaticity distortion, *IEEE Transactions on Consumer Electronics* 53 (3) (2007) 1177–1185.
- [31] T. Horprasert, D. Harwood, L.S. Davis, A statistical approach for real-time robust background subtraction and shadow detection, in: *Proceedings of the IEEE International Conference on Computer Vision*, Kerkyra, Greece, 1999, pp. 1–19.
- [32] K.E. Hoff, J. Keyser, M. Lin, D. Manocha, T. Culver, Fast computation of generalized Voronoi diagrams using graphics hardware, in: *Proceedings of the International Conference on Computer Graphics and Interactive Techniques*, Los Angeles, CA, 1999, pp. 277–286.
- [33] M. Kazhdan, M. Bolitho, H. Hoppe, Poisson surface reconstruction, in: *Symposium on Geometry Processing*, Sardinia, Italy, 2006, pp. 61–70.
- [34] H. Kato, ARToolKit. Available from: <<http://www.hitl.washington.edu/artoolkit/>>.
- [35] C. Liang, K.-Y.K. Wong, Exact visual hull from marching cubes, in: *Proceedings, International Conference on Computer Vision Theory and Applications*, 2008, pp. 597–604.
- [36] A. Erol, G. Bebis, R.D. Boyle, M. Nicolescu, Visual hull reconstruction using adaptive sampling, in: *Proceedings of the IEEE Workshops on Application of Computer Vision*, 2005, pp. 234–241.
- [37] Yebin Liu, Qionghai Dai, Wenli Xu, Free viewpoint video data sets. Available from: <<http://media.au.tsinghua.edu.cn/fvv.jsp>>.
- [38] Y. Yemez, F. Schmitt, Multilevel representation and transmission of real objects with progressive octree particles, *IEEE Transactions on Visualization and Computer Graphics* 9 (4) (2003) 551–569.