Improved space carving method for merging and interpolating multiple range images using information of light sources of active stereo

Ryo Furukawa[1], Tomoya Itano[1], Akihiko Morisaka[1], and Hiroshi Kawasaki[2]

[1] Faculty of Information Sciences, Hiroshima City University,
3-4-1, Ozuka-higashi, Asaminami-ku, Hiroshima, Japan
`ryo-f@cs.hiroshima-cu.ac.jp`
`{t_itano,a_morisaka}@toc.cs.hiroshima-cu.ac.jp`
[2] Faculty of Engineering, Saitama University,
255, Shimo-okubo, Sakura-ku, Saitama, Japan
`kawasaki@cgv.ics.saitama-u.ac.jp`

Abstract. To merge multiple range data obtained by range scanners, filling holes caused by unmeasured regions, the space carving method is a simple and effective method. However, this method often fails if the number of the input range images is small, because unseen voxels that are not carved out remains in the volume area. In this paper, we propose an improved algorithm of the space carving method that produces stable results. In the proposed method, a discriminant function defined on volume space is used to estimate whether each voxel is inside or outside of the objects. Also, in particular case that the range images are obtained by active stereo method, the information of the positions of the light sources can be used to improve the accuracy of the results.

1 Introduction

Shape data obtained by 3D measurement systems are often represented as range images. In order to obtain the entire shape model from multiple scanning results, it is necessary to merge these range images. For this task, several types of approaches have been proposed. Among them, the methods using signed distance fields have been widely researched because they are capable of processing a large volume of input data efficiently. Signed distance fields have also been used as a shape representation in order to interpolate holes appearing in unmeasured parts of object surfaces. Curless and Levoy filled holes of a measured shape by classifying each voxel in volume space as either Unseen (unseen regions), Nearsurface (near the surfaces), or Empty (outside the object), and generating a mesh between Unseen and Empty regions [1]. This process is known as space carving (SC) method. The SC method is capable of effectively interpolating missing parts when there is plenty of observed data. However, when only a few range images are captured, the SC method often fails. This problem occurs because the target object and the ”remains of carving” in volume space become connected.

One of the approaches to solve this problem would be classifying the Unseen voxels as either inside or outside of the object. In this paper, object merging and the interpolation method is proposed based on this approach. Since Unseen voxels include both unobserved voxels inside an object (due to occlusion or low reflection) and voxels outside the object, it is necessary to discriminate these cases. To classify voxels, we take the following two approaches: (1) defining discriminant function for classifying Unseen voxels, and (2) using the positions of light sources if the range images are obtained using active stereo methods. Under the proposed method, the large “remains of carving” that often occurs in SC method are not generated. Also, since all voxels are classified as inside, outside or near the surface, closed shapes are always obtained.

A unique property of the proposed algorithm is that it can be implemented on GPU. Recently, many methods have been proposed for utilizing the computational performance of GPUs for general calculations besides graphics. Our algorithm can be executed efficiently on GPU.

2 Related work

To merge multiple range data into a single shape, several types of approaches have been proposed, such as, for example, generating a mesh from unorganized points [2], using deformable model represented as a level-set [3], stitching meshes at the overlapped surfaces (zippering)[4], and methods using signed distance fields [1,5,6]. In particular, methods using signed distance fields have been widely researched since they are capable of processing a large volume of input data. In order to express the distance from a voxel to the object’s surface, the Euclidean distance (the shortest distance between the center of the voxel and the object’s surface) is theoretically preferable[5,6]. However, since computational cost of calculating the Euclidean distance is large, simplified methods such as line-of-sight distance (the distance between the voxel center and the object’s surface measured along the line of sight from the viewpoint of the measurement) are sometimes used [1] instead.

Regarding filling holes of the surface of the shape defined as isosurfaces of signed distance fields, a number of methods have already been proposed. Davis et al. [7] presented a method in which signed distance field volume data is diffused using a smoothing filter. As shown in the experiments later, this method sometimes propagates the isosurface in incorrect directions and yields undesirable results (Figure 4(b)).

According to the SC method proposed by Curless et al. [1], all voxels are first initialized as Unseen. Then, all voxels between the viewpoints and the object surfaces are set to Empty (this method carves the voxels in front of the observed surfaces, in contrast to the SC method used for 3D reconstruction from 2D silhouettes in which the voxels in the surrounding unobserved regions are carved). This method only carves the volume space in front of the observed surfaces, so, in practice, the unobserved voxels outside of the object remain as Unseen, and excess meshes are generated on the borders between the Unseen and the Empty regions. When range images from a sufficient number of observation points are merged, this kind of excess meshes are not connected to the target object mesh, and can be simply pruned away. However, when the number of observation points is small, or when the object is not observed from certain directions, the excess meshes often become connected to the object (Figure 4(a)).

Sagawa et al. succeeded in interpolating scenes with complex objects by taking the consensus between each voxel in a signed distance field with its surrounding voxels [8]. Masuda proposed a method for filling unmeasured regions by fitting quadrics to the gaps in the signed distance field [9]. They use a Euclidean distance for their calculation to achieve high quality interpolation at the cost of high computational complexity.

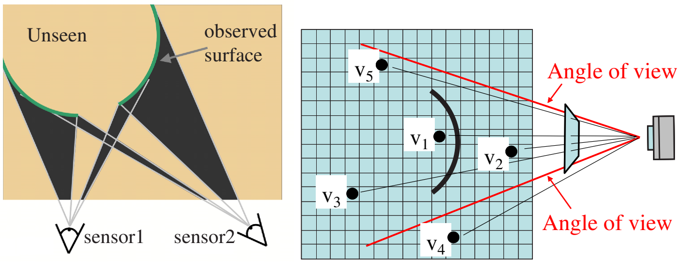


Fig. 1. (Left)Shape interpolation using space carving, and (right) classes of voxels used in the proposed method.

3 Shape merging and interpolation using class estimation of unseen voxels

3.1 Outline

A signed distance field is a scalar field defined for each voxel in 3D space such that the absolute value of the scaler is the distance between the voxel and the object surface and the sign of the scaler indicates whether the voxel is inside or outside of the object (in this paper, voxels inside the object are negative, and those outside are positive). By describing a signed distance field as $D(x)$, it is possible to define an object's surface as the isosurface satisfying $D(x) = 0$. In order to express the distance from a voxel to the object's surface, although there exist several problems for accuracy of hole-filling process, we adopt the line-of-sight distance (the distance from the voxel center to the object's surface measured along the line of sight from the viewpoint of the measurement) instead of the Euclidean distance, since its computational cost is relatively small.

The signed distance $D(v)$ for a voxel v is calculated by accumulating the signed distances from each of the range images, $d_1(v), d_2(v), \cdots, d_n(v)$, each multiplied by a weight $w_i(v)$. It is obtained with the following formula.

$$D(v) = \sum_i d_i(v) w_i(v) \tag{1}$$

The weights represent the accuracy of each distance value, and are often decided according to the angles between the directions of the line-of-sight from the camera and the directions of the normal vectors of the surface.

In the constructed signed distance field $D(x)$, the isosurface satisfying $D(x) = 0$ is the merged shape. In order to generate the mesh model of the merged result, the marching cubes method [10] is used.

The SC method of Curless et al. [1], divides all the voxels in volume space into one of the three types: Unseen (not observed), Empty (outside the object) and NearSurface (near the surface of the object). Shape interpolation is then performed by generating a mesh over the border between Unseen and Empty voxels (Figure 1(left)). A large problem with this method is that all the voxels in the following three cases are classified as Unseen: (1) voxels outside the object, that do not exist on any line-of-sight to an observed region in the range images, (2) voxels outside the object that exist BEHIND a surface of a range image (occluded voxels outside the object), (3) voxels inside the object. In the proposed method, each Unseen voxel is classified as either outside or inside of objects using a discriminant function to solve this problem.

We now describe the classification of voxels according to the proposed method. For a given voxel, the information obtained from a range image takes one of the following four types ((Figure 1(right)).

- **NearSurface (near the surface)**: The absolute value of the signed distance is below the assigned threshold, the voxel in question is classified as "near the surface", and the signed distance is retained (case of v_1 in Figure 1(right)).
- **Outside (outside the object)**: The absolute value of the signed distance is larger than the threshold and the sign is positive. The voxel in question exists between the object and the camera, so the voxel can be classified as "outside the object" (v_2).
- **Occluded (occluded region)**: The absolute value of the signed distance is larger than the threshold and the sign is negative. It is not possible to assert unconditionally whether the voxel is inside or outside the object. In this case, the classification of the voxel in question is temporarily set to Occluded. Whether the voxel is inside or outside is estimated afterwards (v_3).
- **NoData (deficient data)** : The signed distance value cannot be obtained due to missing data in the range images. It cannot be judged whether the voxel is inside or outside. In this case, the classification of the voxel is temporarily set to NoData. Whether the voxel is inside or outside is estimated afterwards (v_5).

The case of v_4 in Figure 1(right), when the voxel in question is outside the angle of view of the range image, may be handled as either Outside or NoData according to the application. For many applications, the voxels outside of the view can be treated as Outside, but in cases such as zooming and measuring a large object, they should be treated as NoData.

When merging multiple range images, the information obtained from each range image regarding voxels may differ. In such cases, priority is given to NearSurface over other classes. The second priority is given to the Outside class. When the information from range images is only Occluded or NoData, it is estimated whether the voxel is inside or outside of the object according to the discriminant function defined in Section 3.2. The classified voxels are tagged as Inside or Outside.

By performing the above process, all the voxels are classified into the three types: Inside (inside the object), NearSurface (near the surface), and Outside (outside the object). The usual signed distances are assigned to the NearSurface voxels, and a fixed negative and positive values are assigned Inside and Outside voxels, respectively. By generating the isosurface of the constructed signed distance field, a merged polygon mesh can be obtained.

3.2 Voxel class estimation

If the information of a voxel obtained from all the range images is only Occluded or NoData, whether the voxel is inside or outside the object should be estimated. For this purpose, discriminant functions based on Bayes estimation are defined.

We consider scalar values that are positive outside the object and negative inside the object, and we estimate the probability distributions of these values based on Bayes estimation. The subjective probability distribution when there is no data is a uniform distribution. Based on the posterior probabilities obtained from the data calculated from each range image, the subjective probability distributions are updated according to Bayes theory. Using normal distributions $N(\mu, \sigma)$ for the posterior distributions, a heuristics for assigning the voxel to Inside or Outside is represented as a mean value μ, and the degree of confidence in the heuristics is represented as the standard deviation σ (the higher σ, the lower the degree of confidence).

When the voxel in question is Occluded in a given range image, the voxel is behind the surface from the viewpoint. When the absolute value of the distance from the surface of the voxel (expressed as $Dist$) is relatively small, the confidence that the voxel is inside the object is high. On the other hand, when it is far from the surface ($Dist$ is large), the confidence that it is inside the object is low. In this case, the degree of confidence that the voxel is inside is $1/Dist$, so the corresponding posterior distribution is $N(-1, Dist)$.

When the voxel in question is NoData for a given range image, it may be either an outside voxel or an unobserved voxel inside the object. For actual cases, pixels of NoData in the range images often indicate outside regions. From this heuristics, a constant value is assigned to the degree of confidence that the voxel is outside, so the posterior distribution corresponding to NoData is represented as $N(1, Const)$, where $Const$ is a user-defined parameter. According to experiments, reasonable results can be obtained by setting $Const$ to a value near the smallest thickness of the object.

In Bayes estimation using a normal distribution, the prior distribution $N(\mu, \sigma)$ is updated using the posterior distribution $N(\mu', \sigma')$ as follows.

$$\mu \leftarrow \frac{\sigma'\mu + \sigma\mu'}{\sigma + \sigma'} \;\; , \;\; \frac{1}{\sigma} \leftarrow \frac{1}{\sigma} + \frac{1}{\sigma'} \tag{2}$$

After performing this for all range images, a voxel is classified as Outside if the sign of the mean value μ of the final probability distribution is positive, and Inside if it is negative. By defining the discriminant function $C(v)$ for voxel v as

$$c_i(v) = \begin{cases} -1/|d_i(v)| & \text{if } Occ(i, v) \\ 1/Const & \text{if } Nod(i, v) \end{cases} , \tag{3}$$

$$C(v) = \sum_i c_i(v), \tag{4}$$

the above judgment is equivalent to determining the inside and outside of the object according to the sign of $C(v)$, where $Occ(i, v)$ and $Nod(i, v)$ mean that the information for the voxel v obtained from the range image i is Occluded or NoData, and $d_i(v)$ is the signed line-of-sight distance between the voxel v and the surface of the range image i.

3.3 Utilizing the position of light sources in active stereo method

Active stereo methods ,in which the light sources are used as well as the camera, are representative examples of techniques for obtaining dense 3D shapes. In these methods, the occlusion of light sources causes missing data, but when data is measured at a point, it proves that the voxels between the point and the light source are outside of the object.

In this research, we utilize the position of light sources in an active stereo system by using additional range images from virtual cameras located at the light sources. For each of the measurement, we consider using two range images, from both the camera and the light source position (described below as the camera range image, and light source range image). The light source range image can be generated by projecting the 3D positions of pixels of the camera range image onto the virtual camera at the light source position.

In the case shown in Figure 2(a), the data ends up missing in the camera range image since the light from the light source is occluded, therefore, at the voxel shown in Figure 2(b), the range data from the camera range image is missing. However, by referring to the light source range image, the line-of sight distance from the light source position to the voxel may be obtained. There are two advantages in using this information. First, by using light source range images, the number of voxels that can be classified as NearSurface or Outside increases. In addition, even if a voxel is not classified to these classes, we can define an improved discriminant function that utilizes more information than the one described in Section 3.2.

The voxel types are mainly the same as when using the camera range images alone, but the order of priority for the voxel classifications is set to NearSurface in the camera range image, NearSurface in the light source range image, Outside in the camera range image, followed by Outside in the light source image. The discriminant function $C(v)$ for voxel classification that utilizes the positions of the light sources is as follows.

$$c_i(v) = \begin{cases} \frac{-1}{|d_i^c(v)|} + \frac{-1}{|d_i^l(v)|} & \text{if } Occ^c(i,v) \wedge Occ^l(i,v) \\ -1/|d_i^c(v)| & \text{if } Occ^c(i,v) \wedge Nod^l(i,v) \\ -1/|d_i^l(v)| & \text{if } Nod^c(i,v) \wedge Occ^l(i,v) \\ 1/Const & \text{if } Nod^c(i,v) \wedge Nod^l(i,v) \end{cases}, \tag{5}$$

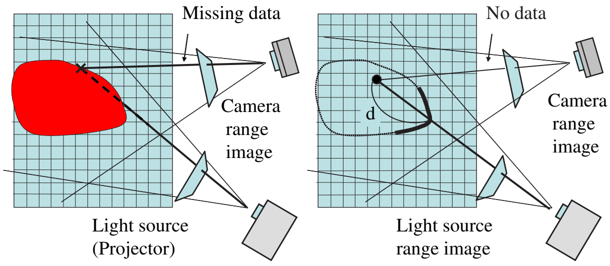
$$C(v) = \sum_i c_i(v). \tag{6}$$

The superscript symbols c and l of $Occ, Nod, d_i(v)$ mean the camera range images and light source range images, respectively.

4 Implementation using a GPU

The merging and interpolation algorithms in this paper can be executed efficiently by utilizing a GPU. Since GPUs can only output 2D images, the volume of the signed distance field is calculated slice by slice. The signed distance value $D(v)$ and discriminant function value $C(v)$ are calculated by rendering each slice in a frame buffer.

Rendering is performed by multi-pass rendering using programmable shaders. Each pass performs the rendering for one range image, and the results are added using blending functions. When performing the rendering for the ith camera range image (range image i), the camera range image and the corresponding light source range image are treated as floating point textures. Then, by issuing an instruction to draw one rectangle



(a) (b)

Fig. 2. Using light sources of active stereo methods: (a) Missing data caused by occlusion, and (b) using light source range images.

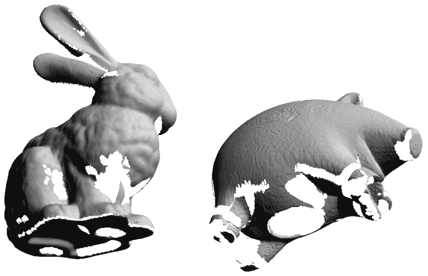
in the whole frame buffer, the pixel-shader process for each pixel is executed. The pixel shaders calculate $d_i(v)w_i(v)$ and $c_i(v)$ using the range image and voxel positions as input values, and add them to the frame buffer. This process is repeated for the number of measured shapes, while changing the textures of the camera range images and the light source range images. Finally, the frame buffer holding the weighted sum of signed distances $D(v)$, and the values of $C(v)$ is read back into the CPU. The flags for voxel classes are checked, the signed distance and the discriminant function are combined, and a slice of the signed distance field is calculated.

Since only small parts in the entire volume space are related to the isosurface, the processing cost can be reduced by first computing the signed distance field in a coarse resolution, and performing the same computation in the high resolution for only the voxels determined to be near the isosurfaces in the coarse resolution. In the experiments described in section 5, the implementation of the proposed method uses this kind of coarse-to-fine method to reduce the computational cost and time.

5 Experiments

In order to demonstrate the effectiveness of the proposed method, experiments were conducted using two types of data: synthetic data formed by seven range images generated from an existing shape model (mesh data), and an actual object (an ornament in the shape of a boar) measured from 8 viewpoints with an active stereo method. For synthetic data, the points on the surface where the light is self-occluded by the object were treated as missing data even if they were visible from the camera, as occurs in measurements by an active stereo method.

Each data set was merged and a mesh model was generated using the SC method, the volumetric diffusion method [7] (VD method), the method proposed by Sagawa et al. [8] (Consensus method), and the proposed method (the information regarding the light source position for active stereo was used). In the SC method, the VD method, and the proposed method, the signed distance field is calculated using line-of-sight distance. In the Consensus method, however, the signed distance field is calculated using the Euclidean distance. In the VD method, the diffusion of the volume was repeated 30 times. Also, the size of the volume was $512 \times 512 \times 512$. For efficiency, in the proposed method, the procedure is first executed with resolution of $62 \times 62 \times 62$ and rendering



(a)Synthetic data (b)Real data

Fig. 3. Results of merging(without interpolation)

(a)SC (b)VD (c)Consensus (d)Proposed(Smoothing)

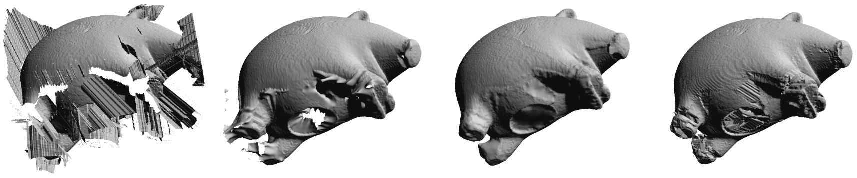
Fig. 4. Results of merging(synthetic data)

with the high resolution was only performed for the region where the surface exists. We performed experiments regarding the proposed method, applying a smoothing filter with a $3 \times 3 \times 3$ size to the volume, and also without such a filter. The filter prevents aliasing on the interpolated surface (the smoothing was not performed in the SC method or the VM method). The SC method, the VD method and the proposed method were executed on PC with an Intel Xeon (2.8GHz) CPU, and an NVIDIA GeForce8800GTX GPU installed. The Consensus method was implemented on a PC constructed with 2 Opteron 275 (2.2GHz) CPUs (a total of 4 CPUs).

The results of merging with no interpolation are shown in Figure 3, and the results of the interpolation process with each method (for the proposed method, the case when smoothing was applied) are shown in Figures 4 and 5. Also, Figures 6(a)-(f) show the signed depth fields for each experiments on the synthetic data sliced at a certain z-coordinate.

Figure 4(a) and Figure 5(a) show that excess meshes were produced surrounding the target object using the SC method, since Unseen (not observed) regions remained uncarved.

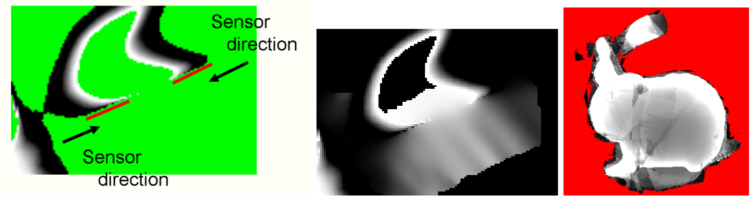
Figure 4(b) shows that, using the VD method, the mesh surrounding the holes beneath the ears and on the chest spread in undesirable directions. These phenomena often occurred where the borders between the NearSurface and Unseen regions (the red line in Figure 6(e)) were not parallel with the normal directions of the surface of the object



(a)SC (b)VD (c)Consensus (d)Proposed(Smoothing)

Fig. 5. Results of merging(real data)

(a)No interpolation (b)SC (c)VD (d)Proposed



(e)Magnified image of (a) (f)Magnified image of (c) (g)Discriminant function

Fig. 6. (a)-(f):Slice images of the signed distance field: For “No interpolation”, “SC” and “VD,” green color represents Unseen, black represent Empty, and gray colors represents signed distance at NearSurface regions. For the proposed method, green represents Inside, black represents Outside. (g):The discriminant function. Gray intensities mean plus values and red color means region of minus value.

as shown in Figure 6(e). In such regions, the expansion of the isosurfaces due to the filter to occur in wrong directions (in Figure 6(f) the isosurface expands downwards and to the right). A similar phenomenon also occurred in the actual data (Figure 5(b)).

The Consensus method produced interpolated meshes with the best quality. However, processing required a long time since the Euclidean distance was used.

The proposed method produced the results whose qualities are the best next to the Consensus method. Since our method does not use signed distance field, but only uses a discriminant function for Unseen voxels to fill holes, smoothness or continuity of the shapes are not considered, thus there remains a good chance to improve the quality. Using both a signed distance field and a discrimination function may be a promising way to improve our algorithm in terms of qualities of results. Figure 6(g) shows the values of the discriminant function $C(v)$. From the figure, we can see that regions where $C(v)$ has plus values coarsely approximate the shapes of the target objects.

Table 1 is the execution times of each method. It shows that the proposed method was executed much faster than the other methods.

Table 1. Execution time in seconds. The volume size of the Consensus method for synthesized data was $128 \times 128 \times 128$.

Methods	SC	VD	Consensus	Proposed	
				No smoothing	Smoothing
Synthetic data	55	168	15 min. for merging, 18 sec. for interpolation	25	36
Real data	38	120	7 hours for merging, 280 sec. for interpolation	21	28

6 Conclusion

In this paper, the space carving method was improved, and an interpolation algorithm yielding stable results even when there are few range images was proposed. This method was realized by defining a discriminant function based on Bayes estimation in order to determine the inside and outside of an object in a signed distance field, even for unseen voxels. In addition, a method was proposed for improving the accuracy of this discriminant function by using range images obtained using an active stereo method. Furthermore, a technique for implementing the proposed method using a GPU was stated, which realized a reduction in computational time by a wide margin. Finally, experiments were conducted regarding the proposed method and existing methods, and the effectiveness of the proposed method was confirmed.

Acknowledgement

This work was supported in part by SCOPE(Ministry of Internal Affairs and Communications, Japan).

References

1. Curless, B., Levoy, M.: A volumetric method for building complex models from range images. Computer Graphics **30**(Annual Conference Series) (1996) 303–312 1, 2, 3
2. Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., Stuetzle, W.: Surface reconstruction from unorganized points. In: ACM SIGGRAPH, ACM Press (1992) 71–78 2
3. Whitaker, R.T.: A level-set approach to 3d reconstruction from range data. IJCV **29**(3) (1998) 203–231 2
4. Turk, G., Levoy, M.: Zippered polygon meshes from range images. In: SIGGRAPH '94, ACM Press (1994) 311–318 2
5. Masuda, T.: Registration and integration of multiple range images by matching signed distance fields for object shape modeling. CVIU **87**(1-3) (2002) 51–65 2
6. Sagawa, R., Nishino, K., Ikeuchi, K.: Adaptively merging large-scale range data with reflectance properties. IEEE Trans. on PAMI **27**(3) (2005) 392–405 2
7. Davis, J., Marschner, S.R., Garr, M., Levoy, M.: Filling holes in complex surfaces using volumetric diffusion. In: Proc. of International Symposium on the 3D Data Processing, Visualization, and Transmission (3DPVT '02). (2002) 428–438 2, 7
8. Sagawa, R., Ikeuchi, K.: Taking consensus of signed distance field for complementing unobservable surface. In: Proc. 3DIM 2003. (2003) 410–417 2, 7
9. Masuda, T.: Filling the signed distance field by fitting local quadrics. In: Proc. of International Symposium on the 3D Data Processing, Visualization, and Transmission (3DPVT '04), Washington, DC, USA, IEEE Computer Society (2004) 1003–1010 2
10. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. In: SIGGRAPH '87, New York, NY, USA, ACM Press (1987) 163–169 3