Lighting Compensating Multiview Stereo

Dimitrios Chrysostomou, Lazaros Nalpantidis and Antonios Gasteratos

Laboratory of Robotics and Automation, Department of Production and Management Engineering
Democritus University of Thrace, PME Building I, Vas. Sophias 12, GR-671 00 Xanthi, Greece
Email: dimitris.chrysostomou@ieee.org, lanalpa@ieee.org, gasteratos@ieee.org

***Abstract*—In this paper, a method that performs 3D object reconstruction from multiple views of the same scene is presented. This reconstruction method initially produces a basic model, based on the space carving algorithm, that is further refined in a subsequent step. The algorithm is fast, computationally simple and produces accurate representations of the input scenes. In addition, compared to previously presented works the proposed algorithm is able to cope with non uniformly lighted scenes due to the characteristics of the used voxel dissimilarity measure. The proposed algorithm is assessed and the experimental results are presented and discussed.**

I. Introduction

Contemporary complex mechatronic systems often have to deal with objects in 3D space. Thus, a 3D representation of these objects is demanded, instead of raw sensorial input, for further processing. In this work we consider the classical problem of inferring a dense 3D structure reconstruction of an object from a collection of views calibrated to a common world coordinate system. Among the multitude of existing methods one can distinguish the use of methods ranging from stereo vision-based to volumetric approaches. In stereo vision [1], the aim is to recover the correspondences across the views based on pixel differences so as to use the 3D view geometry to compute a point cloud containing the structure information of the scene. Solving the correspondence problem is a demanding task and has received much attention recently [2], [3], [4]. Several techniques include the usage of bundle adjustment methods [5], [6] to find the intrinsic and extrinsic parameters of the cameras. Once the calibration of the cameras is achieved, a metric reconstruction can be deduced from the image data.

In contrast to bundle adjustment, which is basically the processing of bundles of light rays traveling through the camera centers, other approaches found in literature are based upon volumetric representations. Several approaches have been developed in the last decades: multi-view stereo matching techniques [7], [8], silhouette and stereo fusion methods for 3D object modeling [9], [10], the combination of both multi view techniques and fusion methods on convex domains [11] and several others as mentioned in [12]. Space carving methods are relatively fast and robust enough for recovering 3D information about the structure of an object. In these methods a formation of voxels is projected onto the two-dimensional views of the scene and carved if they are related to the background color. Originally, they have a minimal set of constraints [13] because these methods can fail when dealing with incurvatures and discrepancies.

Other volumetric representations like voxel coloring [14] and the generalized form of it [15] determine whether, or not, to remove a voxel using its visibility properties. In this set of methods, every voxel is visited in a visibility compatible order, where only voxels that have already been checked are permitted to occlude the current voxel. The volumetric space is divided by a sweeping plane moving far away from the camera and every voxel within the plan, is not allowed to occlude another one. Thus, when a voxel is initially examined its visibility in every input image is uniquely determined. Most authors use a photo-consistency metric as a next step in order to decide if the currently examined voxel belongs to the objects volume or should be carved away. A common approach is to apply a global threshold for light variations around the scene. However, using a single threshold is very difficult to achieve favorable results due to several noise and quantization errors. In this manner, some researchers have expressed the three dimensional reconstruction problem as an energy minimization one. Kolmogorov and Zabih in [16] apply a graph cut method to optimize the volume reconstruction in a direct way.

Some other works present three-dimensional reconstruction methods based on probabilistic schemes. Broadhurst in [17] suggests a probabilistic framework for space carving, where each voxel is assigned a probability using a Gaussian model and classified using the Bayes theorem, which is computed by comparing the likelihoods for the voxel to exist and not exist. Zeng in [18] extends previous carving methods for non-Lambertian objects from an arbitrary set of calibrated images using reflectance models and statistics on the photo-consistency criterion. Finally, Bhotika in [19] works on a probabilistic theory based on visibility, occupancy, emptiness and photo-consistency of the Photo Hull distribution.

A. Algorithm Description

The presented method lies within the category of *Space Carving* algorithms. The main concept of the algorithm is to represent the volume of the scene around an object by a grid of voxels and to gradually dissolve this volume carving away successive layers of voxels that have high discrepancy. The block diagram of the presented method is shown in Fig. 1. The scene is captured by different viewpoints, as shown in Fig. 2, in order to obtain a scene's image set.

Initially, the image set is projected into an array of voxels. Then, a simple silhouette segmentation is applied and those voxels related to the image foreground are maintained, while

978-1-61284-896-9/11/$26.00 ©2011 IEEE

Fig. 1. Block diagram of the proposed method.

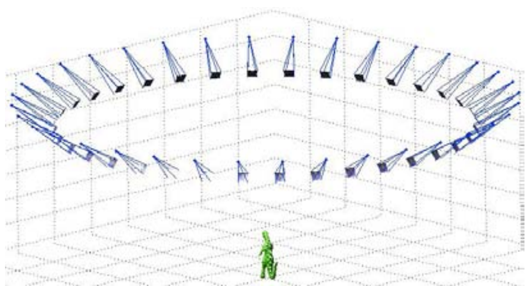


Fig. 2. Positions of cameras capturing a scene

Fig. 3. Input image and extracted silhouette

those related to the background are removed, as shown in Fig. 3. However, space carving difficulties are presented if images are segmented incorrectly, because a voxel could be removed by mistake and then it cannot be retrieved.

Literally, it means that in the final 3D model there could be artificial holes that do not exist in the real input model. Originally, in order to check the existence of a voxel inside the resulting 3D volume the variance of the color samples is needed to be to computed. A large amount of variance specifies that the voxels are not likely to originate from the same surface point and, thus, the specific voxel should be carved away. This is where the main contribution of this paper is. As the viewpoints of the scene can be significantly different, shadowing and other phenomena can differentiate the perceived color of the same surface point. Thus, a lightness compensating similarity (or equivalently dissimilarity) measure is needed.

II. Lightness Compensating Images Comparison

The used dissimilarity measure is defined and calculated within the HSL color space. This color space is represented as a double cone, as is shown in Fig. 4(a). In this figure, H stands for hue and it determines the human impression about which color (red, green, blue, etc) is depicted. Each color is represented by an angular value ranging between 0 and 360 degrees (0 being red, 120 green and 240 blue). S stands for saturation and determines how vivid or gray the particular color seems. Its value ranges from 0 for gray to 1 for fully saturated (pure) colors. The L channel of the HSL color space stands for the Luminosity and it determines the intensity of a specific color. It ranges from 0 for completely dark colors (black) to 1 for fully illuminated colors (white).

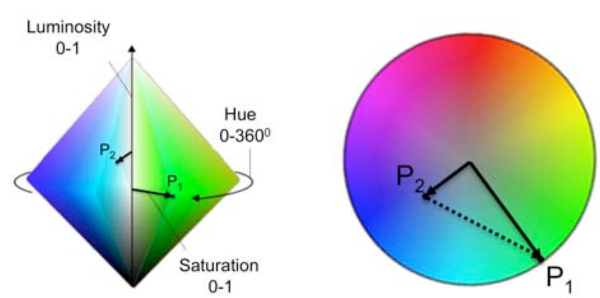
Thus, in the HSL color space lightness is separated by the other characteristics of color. This fact implies that a given color will theoretically result in the same values of hue and saturation regardless the environment's illumination conditions. Ignoring the Luminosity channel will inevitably lead to the loss of some amount of information and, as a result, to slightly inferior results for ideal lighting, but will provide robustness against real, non-ideal and non-uniform lighting conditions [20]. The omission of the vertical (L) axis from the color space representation leads to 2D circular disk, defined only by H and S, as show in Fig. 4(b).

In this reduced colorspace each color $\mathbf{P}_i$ can be represented as a planar vector with its initial point being the disc's center. As a consequence, it can be described as a polar vector or equivalently as a complex number with modulus equal to S_i and argument equal to H_i. That is, a color in the new luminosity-ignoring colorspace representation can be described as:

$$\mathbf{P}_i = S_i e^{iH_i} \tag{1}$$

Based on this color description a luminosity-compensated dissimilarity measure (LCDM) has been proposed in [20]. According to this, the variance of two colors $\mathbf{P}_1$ and $\mathbf{P}_2$ can be found in the reduced HS colorspace of Fig. 4(b) as the difference of the two complex numbers:

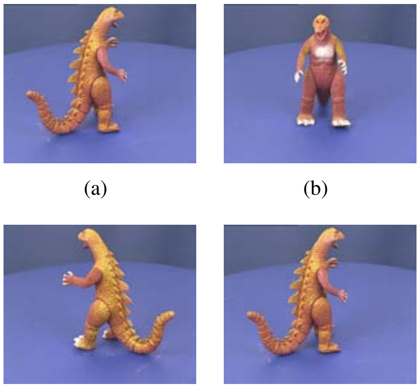
$$\begin{aligned} LCDM_{P_1,P_2} &= |\mathbf{P}_1 - \mathbf{P}_2| \\ &= \left|S_1 e^{iH_1} - S_2 e^{iH_2}\right| \\ &= \sqrt{{S_1}^2 + {S_2}^2 - 2S_1 S_2 \cos(H_1 - H_2)} \end{aligned} \tag{2}$$



(a) The double cone representation

(b) 2D floor plan projection

Fig. 4. Views of the HSL color space representation



(a) (b)

(c) (d)

Fig. 5. Indicative input images from the dinosaur data set (4 out of 36 total)

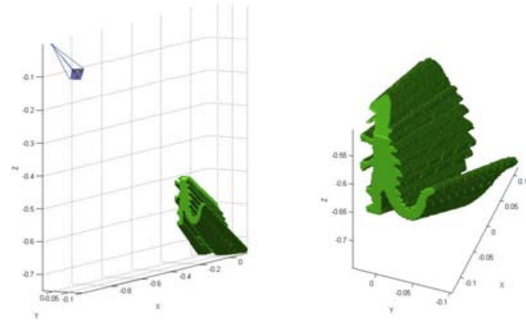
Equation 2 is the mathematical formulation of the LCDM dissimilarity measure, which takes into consideration any chromatic information available, except the luminosity. In contrast to other popular dissimilarity measures, such as absolute differences (AD) or squared differences (SD), LCDM can provide robust behavior against viewpoint-dependent chromatic differentiations. Consequently, it was preferred for use in the presented multiview algorithm.

III. Experimental Results

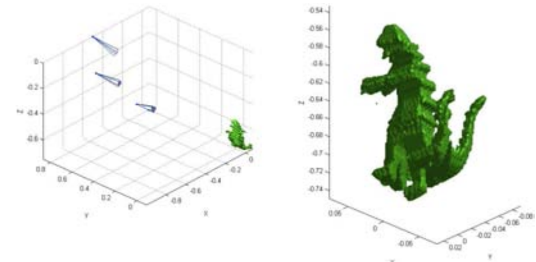
The image set named "Dinosaur" located in the University of Oxford and in the Visual Geometry Group [21] in particular, has been used for testing the preliminary performance of our algorithm. While Fig. 2 shows all the image set used for the object modeling, in Fig. 5 some real images from the toy are presented and the actual viewpoints can been examined. Figure 6(a) shows the initial result after the first carving is applied over the voxel array. It is clear that with just one carving there is no depth presence and more photos are needed in order to obtain reliable three dimensional information about the object. Even when just three carvings are used, as in Fig. 6(b), significant progress can be observed.

Figure 7(a), illustrates the result of the proposed algorithm after all 36 images of the dataset are processed. It can be seen that after the processing of all the image data, most of the uncertainties are eliminated but the final surface is a little harsh. Figure 7(b) depicts the result after the lightness compensation algorithm has been applied and therefore the surface of the model is smoother than the initial result. Finally, in the colored surface of the model in Fig. 7(c) the lack of uncertain voxels and the final refinement of the object is shown.

The proposed space carving algorithm have been implemented in Matlab using a common computer equipped with a Quad Core processor synchronized at 2.83Mhz for each core and with 4GB of DDR3 RAM. The algorithm is not yet optimized to process the voxels with a compact way, such as octrees and utilizes a brute force method. Nevertheless, the relationship between the number of voxels processed and the respecting computational times are apposed in Table I. It can be noticed that the algorithm is relatively fast and produce remarkable results.



(a) Result after one carving



(b) Result after three carvings

Fig. 6. Gradual generation of the 3D model

IV. Conclusion

A method that performs 3D object reconstruction from multiple views of the same scene has been presented. The method is based on a space carving algorithm equipped with a lighting compensating dissimilarity measure. The preliminary results of the used dissimilarity measure exhibits robust behavior against lightness differentiations, which is very important for multi view algorithms. The overall algorithm is computationally efficient and produces an accurate 3D model of the input scenes. More multi view datasets are going to be examined for checking thoroughly the robustness of the algorithm and a comparison with other multi view stereo methods will be implemented to present the qualitative results of the examined approaches. The algorithm's structure is inherently simple and,

(a) Initial result (b) Result after photo-consistency check (c) Final refined result

Fig. 7. Reconstruction results for the dinosaur data set

TABLE I
COMPUTATION TIMES FOR VARIOUS VOXEL POPULATIONS

Number of voxels	**Computation time (sec)**
62450	46
70500	52
81300	57
104510	62
111732	67
123247	73
140241	86
145572	89

thus, can be further accelerated by being implemented on programmable graphic processing units (GPU). Contemporary solutions, such as the CUDA software, comprise an efficient way of realizing such implementations. The proposed technique is applicable to both sophisticated mechatronic systems, such as painting, grasping and milling ones, as well as to standard robotic and machine vision applications.

REFERENCES

[1] D. A. Forsyth and J. Ponce, *Computer Vision: A modern Approach*, us ed. Upper Saddle River, NJ, USA: Prentice Hall Professional Technical Reference, August 2002.

[2] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[3] T. Lemaire, C. Berger, I.-K. Jung, and S. Lacroix, "Vision-based slam: Stereo and monocular approaches," *International Journal of Computer Vision*, vol. 74, no. 3, pp. 343–364, 2007.

[4] L. Nalpantidis, G. C. Sirakoulis, and A. Gasteratos, "Review of stereo vision algorithms: from software to hardware," *International Journal of Optomechatronics*, vol. 2, no. 4, pp. 435–462, 2008.

[5] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment - a modern synthesis," in *ICCV '99: Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*, ser. Lecture Notes in Computer Science, B. Triggs, A. Zisserman, and R. Szeliski, Eds., vol. 1883. London, UK: Springer-Verlag, September 2000, pp. 298–372.

[6] M. I. A. Lourakis and A. A. Argyros, "Sba: A software package for generic sparse bundle adjustment," *ACM Trans. Math. Softw.*, vol. 36, no. 1, pp. 1–30, 2009.

[7] H. Saito, S. Baba, and T. Kanade, "Appearance-based virtual view generation from multicamera videos captured in the 3-d room," *IEEE Transactions on Multimedia*, vol. 5, no. 3, pp. 303–316, 2003.

[8] G. Vogiatzis, C. Hernández, P. H. S. Torr, and R. Cipolla, "Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2241–2246, June 2007.

[9] C. Hernández and F. Schmitt, "Silhouette and stereo fusion for 3d object modeling," *Computer Vision and Image Understanding*, vol. 96, no. 3, pp. 367–392, 2004.

[10] Y. Yemez and C. J. Wetherilt, "A volumetric fusion technique for surface reconstruction from silhouettes and range data," *Computer Vision and Image Understanding*, vol. 105, no. 1, pp. 30–41, 2007.

[11] K. Kolev and D. Cremers, "Integration of multiview stereo and silhouettes via convex functionals on convex domains," in *ECCV'08 : Proceedings of the 10th European Conference on Computer Vision*, ser. Lecture Notes in Computer Science, D. A. Forsyth, P. H. S. Torr, and A. Zisserman, Eds., vol. 5302. Heidelberg: Springer Berlin, October 12-18 2008, pp. 752–765.

[12] G. Slabaugh, W. Culbertson, T. Malzbender, M. Stevens, and R. Schafer, "Methods for volumetric reconstruction of visual scenes," *International Journal of Computer Vision*, vol. 57, no. 3, pp. 179–199, May 2004.

[13] K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," *International Journal of Computer Vision*, vol. 38, no. 3, pp. 199–218, 2000.

[14] S. Seitz and C. Dyer, "Photorealistic scene reconstruction by voxel coloring," *International Journal of Computer Vision*, vol. 25, no. 1, pp. 1067–1073, November 1999.

[15] W. B. Culbertson, T. Malzbender, and G. G. Slabaugh, "Generalized voxel coloring," in *ICCV '99: Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*, ser. Lecture Notes in Computer Science, B. Triggs, A. Zisserman, and R. Szeliski, Eds., vol. 1883. London, UK: Springer-Verlag, September 2000, pp. 100–115.

[16] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," in *ECCV'02 : Proceedings of the 7th European Conference on Computer Vision*, ser. Lecture Notes in Computer Science. London, UK: Springer-Verlag, 2002, pp. 82–96.

[17] A. Broadhurst, T. Drummond, and R. Cipolla, "A probabilistic framework for space carving," in *ICCV'01: Proceedings of the 8th IEEE International Conference on Computer Vision*, vol. 1. Los Alamitos, CA, USA: IEEE Computer Society, 2001, pp. 388–393.

[18] G. Zeng, S. Paris, and L. Quan, "Robust carving for non-lambertian objects," *International Conference on Pattern Recognition*, vol. 3, pp. 119–122, 2004.

[19] R. Bhotika, D. J. Fleet, and K. N. Kutulakos, "A probabilistic theory of occupancy and emptiness," in *ECCV'02 : Proceedings of the 7th European Conference on Computer Vision*, ser. Lecture Notes in Computer Science. London, UK: Springer-Verlag, 2002, pp. 112–132.

[20] L. Nalpantidis and A. Gasteratos, "Stereo vision for robotic applications in the presence of non-ideal lighting conditions," *Image and Vision Computing*, vol. 28, no. 6, pp. 940 – 951, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/B6V09-4XT3HK8-2/2/1eb67f2e03e6b202a9b2c2141a757514

[21] V. G. Group, "http://www.robots.ox.ac.uk/ vgg/data/data-mview.html."