2D to 3D conversion algorithms

Conference Paper · November 2014

DOI: 10.13140/2.1.3346.7202

1 author:

Miroslav Galabov
Veliko Tarnovo University
17 PUBLICATIONS **16** CITATIONS

SEE PROFILE

2D to 3D conversion algorithms

Miroslav Galabov
Computer Systems and Technologies Department
St. Cyril and St. Methodius University
Veliko Turnovo, Bulgaria
lexcom@abv.bg

Abstract—**According to the depth cues on which the algorithms reply, the algorithms are classified into the following categories: binocular disparity, motion, defocus, focus, silhouette, atmosphere scattering, shading, linear perspective, patterned texture, symmetric patterns, occlusion (curvature, simple transform) and statistical patterns. The article describes and analyzes algorithm that use a single depth cue "statistical patterns".**

Keywords- 2D to 3D conversion, depth cue, depth map.

I. INTRODUCTION

Depending on the number of input images, we can categorize the existing conversion algorithms into two groups: algorithms based on two or more images and algorithms based on a single still image. In the first case, the two or more input images could be taken either by multiple fixed cameras located at different viewing angles or by a single camera with moving objects in the scenes. We call the depth cues used by the first group the multi-ocular depth cues. The second group of depth cues operates on a single still image, and they are referred to as the monocular depth cues. The Table 1 summarizes the depth cues used in 2D to 3D conversion algorithms and their representative works.

TABLE I. DEPTH CUES AND THEIR REPRESENTATIVE ALGORITHMS

The Number of Input Images	***Depth Cues***	***Representative Works***
Two or More Images (binocular or multi-ocular)	Binocular disparity	Correlation-based, feature-based correspondence; triangulation [2][3]
	Motion	Optical flow [2]; Factorization [9]; Kalman filter [10]
	Defocus	Local image decomposition using the Hermite polynomial basis [4]; Inverse filtering [11]; S-Transform [12]
	Focus	A set of images of different focus level and sharpness estimation [5]
	Silhouette	Voxel-based and deformable mesh model [6]
One single image (monocular)	Defocus	Second Gaussian derivative [7]
	Linear perspective	Vanishing line detection and gradient plane assignment [8]
	Atmosphere Scattering	Light scattering model [13]
	Shading	Energy minimization [14]
	Patterned texture (Incorporates relative size)	Frontal texel [15]
	Symmetric patterns	Combination of photometric and geometric constraints [16]
	Occlusion	
	- Curvature - Single Transform	Smoothing curvature and isophote [17] Shortest path [18]
	Statistical patterns	Color-based heuristics [8], Statistical estimators [19]

II. PATTERN RECOGNITION IN DEPTH ESTIMATION

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the A4 paper size.

It can be observed that a lot of the 2D to 3D conversion algorithms are still in the research phase. They are not yet ready for real-time use due to factors such as their high complexity or unsatisfactory quality. As well as improving the existing algorithms, a new trend in this field is to analyze the semantic content of the image and use this knowledge to help reconstruct the 3D object. The depth cue "statistical patterns" plays the central part in this trend. The algorithm that we propose operates on a single color image. No a priori knowledge about the image content is needed. It is also claimed to be fully unsupervised and suitable for real-time applications.

Eight steps are involved in this algorithm. Throughout the process, two intermediate depth maps are constructed, the qualitative depth map and the geometric depth map. In the end, these two depth maps are combined together to generate the final depth map.

The following paragraphs expound on these eight steps.

1) Color-based segmentation

Color-based segmentation identifies the chromatic homogeneous regions present in the image. The image is under-segmented so that main chromatic regions are retrieved and fine details are filtered out.

2) Rule-based regions diction to find specific areas

The segmented image in the RGB color model is converted to the HSI color model. The HSI model is more suitable for color description. Subsequently, the intensity values of the R, G, B, H and S components of each pixel in the image undergo various checks based on a set of color-based rules, which has been learned heuristically. These color-based rules are able to identify six semantic regions possibly present in the image: Sky, Farthest Mountain, Far Mountain, Near Mountain, Land and Other.

3) Qualitative depth map construction

Each semantic region is assigned a depth level, which corresponds to a certain gray level following the trend:

Gray(Sky) < Gray(Furthest Mountain) < Gray(Far Mountain) < Gray(Near Mountain) < Gray(Land) < Gray(Other).

The resultant image is termed the qualitative depth map (Figure 1).

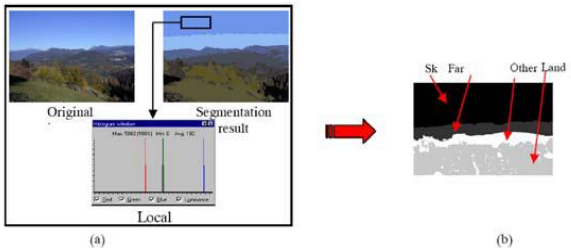

Figure 1. Qualitative depth map

4) Image classification to discriminate three categories: Indoor, Outdoor without geometric appearance, Outdoor with geometric appearance.

The six semantic regions obtain their own labels in this step, for example, the sky is labeled as 's'. The qualitative depth map is then sampled column-wise. Each column is represented by a label sequence which labels from top to down each region present in the column. A typical sequence could be "sml", for instance, which indicates that the sample column consists of the 3 regions -Sky, Mountain and Land (Figure 2).

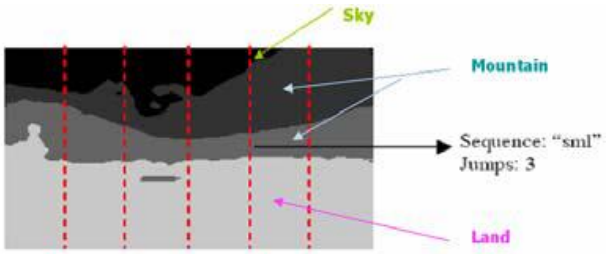

Figure 2. Sampling the qualitative depth map column-wise

After all the sequences in the image have been generated, they are plugged into a counting process to obtain the number of accepted sequences (see Figure 3).

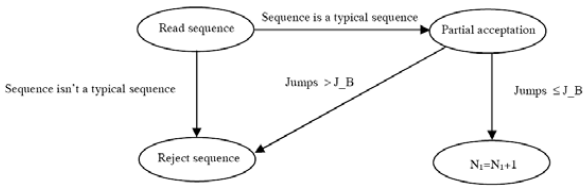

Figure 3. Determining the number of accepted sequences in an image

A sequence is accepted if it belongs to a typical landscape sequence and the number of jumps is smaller than a certain threshold (denoted by J_B in Figure 3). Finally, the following heuristics is applied to classify the image into three categories:

1) #Accepted_sequences > Threshold × #Sequence
⟶ OUTDOOR

2) #Sequences with the first region SKY > Threshold × #Sequences
⟶ OUTDOOR WITH GEOMETRIC APPEARANCE

3) Otherwise ⟶ INDOOR

5) Vanishing lines detection

Different vanishing line detection strategies are applied according to the category to which the image belongs. For Outdoor scenes, the vanishing point is put in the center region of the image and a set of vanishing lines passing through the vanishing points are generated. For the categories Indoor and Outdoor with geometric appearance, a more complex technique is applied (Figure 4). Edge detection (using Sobel operators) and line detection (Hough transform) are conducted to determine the main straight lines. The vanishing point is chosen as the intersection point with the most intersections around it while the vanishing lines are the predominant lines passing close to the vanishing point.

Figure 4. Left: Input image classified as OUTDOOR WITH GEOMETRIC APPEARANCE; Right: Result of vanishing line detection

6) Geometric depth map construction

Taking the position of the vanishing point into account, a set of horizontal or vertical gradient planes is assigned to each neighboring pair of vanishing lines. A gradient plane has a fixed depth level. There are more gradient planes close to the vanishing point than further away because human vision is more sensitive for the depth perception of objects close by.

7) Consistency verification of detected regions

In this step, the qualitative depth map is checked for consistency. False classified semantic regions are detected and corrected. For example, if between two "Sky" regions, there is a region of another type (e.g. a Mountain) with a vertical size larger than certain threshold, the second sky region is then identified as a "false" Sky and its type is changed to the same type of the upper zone.

8) Fusion of the qualitative depth map and the geometric depth map

The final depth map of INDOOR category image is just the geometric depth map. No fusion occurs. For OUTDOOR WITHOUT GEOMETRIC APPEARANCE, the final depth of pixel (x, y) is assigned the depth value in the qualitative depth map in all cases, except when it belongs to the Land ***and*** Other category. In the latter case, pixel (x, y) obtains its depth value from the geometric depth map (see Figure 5). For the image category OUTDOOR WITH GEOMETRIC APPEARANCE, the final depth of pixel (x, y) is assigned the depth value in the geometric depth map for all cases, except when it is a *Sky*

pixel, it then adopts the depth value in the qualitative depth map.

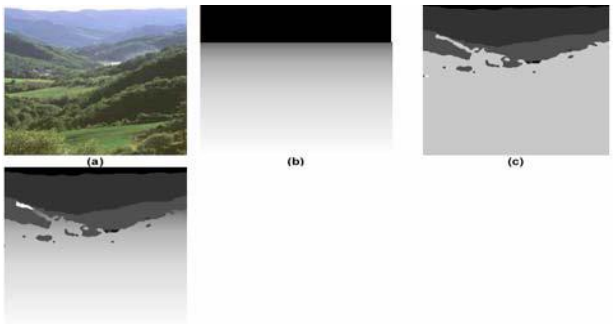

Figure 5. Depth fusion for OUTDOOR WITHOUT GEOMETRIC APPEARANCE

III. CONCLUSIONS

The results of some 2D to 3D conversion algorithms are 3D coordinates of a small set of points in the images. This group of algorithms is less suitable for the 3D television application. The depth cues based on multiple images yield in general more accurate results, while the depth cues based on single still image are more versatile. A single solution to convert the entire class of 2D images to 3D models does not exist. Combing depth cues enhances the accuracy of the results. It has been observed that machine learning is a new and promising research direction in 2D to 3D conversion. And it is also helpful to explore the alternatives than to confine ourselves only in the conventional methods based on depth maps.

ACKNOWLEDGMENT

The presented article is part of research work carried out in the "Analysis, research and creation of multimedia tools and scenarios for e-learning" project - Contract No: RD - 09-590-12/10.04.2013.

REFERENCES

[1] IJsselsteijn, W.A.; Seuntiëns, P.J.H.; Meesters, L.M.J. "State-of-the-art in Human Factors and Quality Issues of Stereoscopic Broadcast Television", Deliverable ATTEST/WP5/01, Eindhoven University of Technology, the Netherlands,2002. URL: http://www.hitech-projects.com/euprojects/attest/deliverables/Attest-D01.pdf

[2] Michel B., "La conversion 2D–3D", in *La Stéréoscopie Numérique,* Eyrolles, Chapter 5, 2011.

[3] Scharstein, D.; Szeliski, R "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms", International Journal of Computer Vision 47(1/2/3), 2002,7-42.

[4] DA SILVA V., "Depth image based stereoscopic view rendering for MATLAB", available at http://www.mathworks.com/matlabcentral/fileexchange/27538-depth-image-based-stereoscopic-view-rendering, 2010.

[5] Guan-Ming Su, Yu-Chi Lai, Andres Kwasinski and Haohong Wang. "3D Visual Communications", First Edition, Published by John Wiley & Sons, Ltd., 2013.

[6] Matsuyama, T. "Exploitation of 3D video technologies ", Informatics Research for Development of Knowledge Society Infrastructure, ICKS 2004, International Conference, 2004, Page(s) 7-14.

[7] Wong, K.T.; Ernst, F., Master thesis "Single Image Depth-from-Defocus", Delft university of Technology & Philips Natlab Research, Eindhoven, The Netherlands, 2004.

[8] Battiato, S. ; Curti, S. ; La Cascia, M.; Tortora, M.; Scordato, E. "Depth map generation by image classification", SPIE Proc. Vol 5302, EI2004 conference 'Threedimensional image capture and applications VI',2004.

[9] Han, M; Kanade, T. "Multiple Motion Scene Reconstruction with Uncalibrated Cameras", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 25, Issue 7, 2003, Page(s): 884 – 894

[10] Franke, U.; Rabe, C., "Kalman filter based depth from motion with fast convergence", Intelligent Vehicles Symposium, Proceedings. IEEE, 2005, Page(s): 181 – 186.

[11] Kao, M.A., Shen, T.C. "A novel real time 2D to 3D conversion technique using depth based rendering". *IDW'09*, p. 203, 2009.

[12] Bleyer M., Gelautz M., "Temporally consistent disparity maps from uncalibrated stereo videos", Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis (ISPA), Salzburg, pp. 383–387, 16–18 September 2009.

[13] XU F., LAM K.M., DAI Q., "Video-object segmentation and 3D-trajectory estimation for monocular video sequences", *Image and Vision Computing Journal*, vol. 29, no. 2–3, pp. 190–205, 2011.

[14] Kang, G; Gan, C.; Ren, W., "Shape from Shading Based on Finite-Element", Proceedings, International Conference on Machine Learning and Cybernetics, Volume 8, 2005, Page(s): 5165 – 5169

[15] Loh, A.M.; Hartley, R., "Shape from Non-Homogeneous, Non-Stationary, Anisotropic, Perspective texture", Proceedings, the British Machine Vision Conference, 2005.

[16] Shimshoni, I.; Moses, Y.; Lindenbaumlpr, M.,"Shape reconstruction of 3D bilaterally symmetric surfaces", Proceedings, International Conference on Image Analysis and Processing, 1999, Page(s): 76 – 81

[17] Redert, A., Patent ID: WO2005091221 A1, "Creating a Depth Map", Royal Philips Electronics, the Netherlands, 2005.

[18] Redert, A.,Patent ID: WO2005083630 A2, WO2005083630 A2, WO2005083631 A2, "Creating a Depth Map", Royal Philips Electronics, the Netherlands, 2005.

[19] Saxena, A; Chung, S.H.; Ng, A. Y., "Learning Depth from Single Monocular Images", Proceedings, 19th Annual Conference on Neural Information Processing Systems, 2005.

[20] Rotem, E.; Wolowelsky, K.; Pelz, D.,"Automatic Video to Stereoscopic Video Conversion", SPIE Proc. Vol. 5664, Stereoscopic Displays and Virtual Reality Systems, 2005.