3D Reconstruction of Indoor and Outdoor Building Scenes from a Single Image

Georgios Vouzounaras[a], Juan Diego Perez-Moneo Agapito[b], Petros Daras[b], Michael G. Strintzis[a,b]

[a]Information Processing Laboratory
Department of Electrical and Computer Engineering
Aristotle University of Thessaloniki
GR54124, Thessaloniki, Greece
+302310996396

[b]Informatics and Telematics Institute
Centre for Research and Technology Hellas
GR57001, Thessaloniki, Greece
+302310464160

gvouzoun@auth.gr, perez@iti.gr, daras@iti.gr, strintzi@eng.auth.gr

ABSTRACT

In this paper, a novel method is proposed able to automatically generate accurate 3D models of both outdoor buildings and indoor scenes with perspective cues from line segments that are automatically extracted from a single image with an uncalibrated camera. The proposed method uses geometric constraints and knowledge of photography and achieves an accurate, real-time and fully automated 3D reconstruction of the scene without any intervention from the user.

Categories and Subject Descriptors

I.4.5 [Reconstruction]

General Terms: Algorithms.

Keywords: Single-view 3D reconstruction, projective correction, vanishing point detection.

1. INTRODUCTION

Image-based modeling relies on a set of techniques for creating a 3D representation of a scene from one or more 2D images. Although humans easily infer depth from images, computer and robots using low-level image processing perform more poorly. Thus, automatic recognition of structure from a collection of line segments is challenging, as not all lines defining the building structure are perfectly detected. Using geometric constraints, such as the parallelism and orthogonality of lines, and the way one building (either its exterior or interior) is structured, we can interpret the collection of the line segments.

Many researchers have dealt with the problem of 3D reconstruction either from a single image or from two or more images, which is more common. The majority of the researchers have managed to solve this problem using a calibrated camera [4] or a partially calibrated camera (a camera that is set to a known height [9]). However, the internal and external parameters of a camera are not always available. Three dimensional information can be extracted from a single image when there is a reference in the image [2]. A commonly used reference is the ground plane. Hoiem *et al.* [7] take a two-step approach for recovering 3D structure of outdoor images: 1) they estimate image region orientation (e.g. ground vertical) using statistical methods on image properties, such as color, texture, edge orientation, position in image, etc. 2) "pop-up" vertical regions by "folding" along the crease between ground and vertical regions. Saxena *et al.* [13] have taken a different approach by estimating absolute depth directly from image properties. J. Huang *et al*. [8] have introduced a really simple algorithm for defining the geometry of an indoor image with perspective cues, which, however, does not deal with the problem of projective distortion. Moreover, commercial systems exist that make robust 3D models, however they require a high amount of user interaction, like Google Sketch Up [5] where user has to define points and lines so as to compute the geometry of the model.

In this paper, 3D geometry is derived from line segments in one-point perspective indoor image that consists largely of orthogonal planes, and from the exterior of buildings. For the indoor scene the presence of the floor, ceiling, and walls in the input images are assumed, at least parts of the floor boundaries are visible, and no window or wall decoration are below the horizon. For the exterior buildings it is assumed that two facades are visible from the photo. The proposed method uses the constraints introduced in [8] and further extends it with removing the perspective distortion so as to make a more accurate 3D model.

The goal of this work is to present a novel algorithm for real-time 3D scene generation from a single image able to be used for both indoor and outdoor images of buildings fully automated, real-time, without any user interaction.

2. PROPOSED ALGORITHM

The proposed algorithm consists of five steps: line segment detection and vanishing point estimation (Section 2.1) in order to render the orientation of the line segments in the image; automatic estimation of the orientation of the image (Section 2.2), if the image is indoor or outdoor; detection of the planes of the image (Section 2.3) so as to rectify the image; image rectification (Section 2.4) depending on each plane that has been detected, in order to extract the 3D points, and 3D reconstruction of the image (Section 2.5).

2.1 Line Segment Detection and Vanishing Point Estimation

Firstly, the canny edge detector is applied on an image to extract edges and the Hough transform is used to link the edges and fit line

SMVC'10, October 29, 2010, Firenze, Italy.

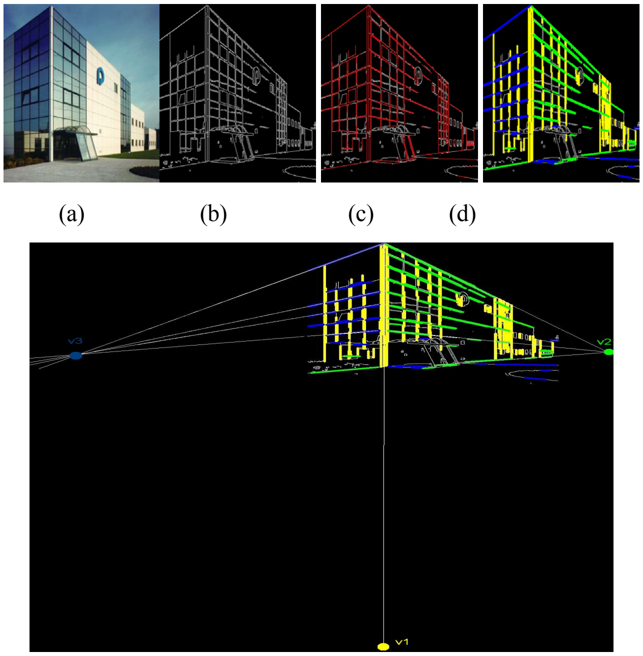
segments. Then, the procedure proposed in [12] is followed so as to identify three orthogonal vanishing points: $v_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix}, i = 1,2,3$

The algorithm in [12] consists of two steps: the first step is the accumulation step, where the intersection points of all pairs of non-collinear line segments are found and the vote of each accumulator cell (intersection point) is determined with a voting equation [12]. The second step is the search step, where the accumulator cell with the highest vote is taken and for all pairs of accumulator cells three criteria should be fulfilled: the camera, orthogonal and vanishing line criterion. Then, the total number of votes of the three accumulator cells is computed and finally, the three vanishing points are the pairs with the highest vote. In this paper, an alteration has been made so as to make the algorithm work so as it can be used in real-time problems. The accumulator cell with the highest vote is being identified, which is the first vanishing point. Then, the line segments that vote for this cell are omitted and the process is repeated. Again, the cell with the highest vote is taken and this is the second vanishing point. Finally, the process is repeated for the last time and the third vanishing point is extracted. The proposed algorithm (Figure 1) is efficient when there are lines to all orthogonal directions, which is obligatory for 3D reconstructions. Figure 2 shows an example of a building and Figure 3 depicts the interior of a building.

```
Algorithm 1 Vanishing point detection

for num=1 to 3
    Find the intersection points of all non-collinear
    line segments.
    for all intersection points
        for all accepted line segments
            compute vote($a_i$) [12]
        end for
    end for
    if num = 1
        find max vote, set v1= max (vote($a_i$))
        extract the lines that vote for this cell
    else if num = 2
        find max vote, set v2= max (vote($a_i$))
        extract the lines that vote for this cell
    else
        find max vote, set v3= max (vote($a_i$))
        extract the lines that vote for this cell
    end if
end for
```

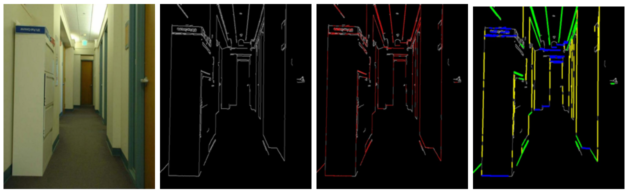
Figure 1. Vanishing point detection algorithm.



(a) (b) (c) (d)

(e)

Figure 2. (a) Original image. (b) Canny edge detection. (c) Hough transform. (d) Lines orientation. (e) Vanishing points.



(a) (b) (c) (d)

Figure 3. (a) Original image. (b) Canny edge detection. (c) Hough transform. (d) Lines orientation.

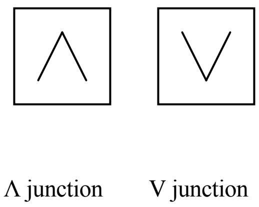
2.2 Image Orientation

After vanishing point detection we can simply decide whether the image is interior or exterior, by checking the vanishing points. If there is only one finite vanishing point and the distance from the center of the image is below a certain threshold t_h $\left\| \mathbf{v}_{1x} - \mathbf{center}_x \right\| \le image(width)/t_h$, $\left\| \mathbf{v}_{1y} - \mathbf{center}_y \right\| \le image(height)/t_h$, then it is interior. t_h was experimentally defined to be 4. If there are two or three finite vanishing points then the image is a building (exterior).

2.3 Planes Detection

2.3.1 Outdoor Image

In the case of the building three planes exist, two facades, the left and the right, and the ground. We find these planes using geometric constraints (Figure 4, Algorithm 2). More specifically, the line that defines these two facades is being found. Below this line is the ground and the image has to be cropped, according to the line, for the rectification process (Section 2.4). The two facades of a building create two types of junctions, as shown in Figure 4. Depending on the junction the proposed algorithm is explained in Figure 4.



(a)

```
Algorithm 2 Planes detection
if $v_2$, $v_3$ out of the image plane
    (like Figure 3) crop line is
    the longest of the vertical
    lines
else if $v_2$, $v_3$ in the image plane
    crop line is the shortest of
    the vertical lines
end if
```

(b)

Figure 4. (a) Two types of junctions. (b) Algorithm of planes detection

2.3.2 Indoor Image

In the case of an indoor image, we find the planes that define the ceiling, floor, walls and occluding objects that exist in the image using the constraints that were proposed in [8].

2.4 Image Rectification

In this step, the perspective distortion of the image is being removed in order to make a precise 3D model [10], [11]. The rectification homography that maps a point **x**, on the image plane to a point **x**′, on the world plane, is represented by a 3x3 homogeneous matrix **H** such that:

$$\mathbf{x}' = \mathbf{H} \cdot \mathbf{x} \tag{2.1}$$

where **x** and **x**′ are homogeneous 3-vectors and **H** has eight degrees of freedom. **H** may be decomposed as:

$$\mathbf{H} = \mathbf{M} \cdot \mathbf{N} \tag{2.2}$$

The transformation **M** is the metric part of the homography and is a similarity transformation, which can be of the form:

$$\mathbf{M} = \begin{pmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \tag{2.3}$$

where **R** is a rotation matrix, **t** a translation vector, and s an isotropic scaling factor. There are four degrees of freedom in **M.** The remaining four degrees of freedom in the rectifying homography are encoded by **N**, the non-metric component of **H**. The established method of computing **H** is from the correspondence of four (or more) points with known position [3], [6]. Given four points (2.1) may be written out explicitly for known **x** and **x**′. The four points give eight equations in the eight degrees of freedom of **H** and the elements of **H** may then be computed. Once **H** is determined the image can be wrapped onto the world plane and in this way a rectified image is obtained.

However, it is not necessary to determine the entire homography in order to obtain a metric rectification. The metric properties of the plane, such as angle and relative length, are invariant to **M** since it is a similarity transformation. The complete metric rectification is thus known when **N** is computed.

The non-metric part of **H** can be decomposed into two matrices:

$$\mathbf{N} = \mathbf{H_a} \cdot \mathbf{H_p} \tag{2.4}$$

The first stage is to determine $\mathbf{H}_p$ which requires identifying the vanishing line $\mathbf{l}_\infty$ of the plane. The vanishing line is the image of the line at infinity on the world plane. Parallel lines on the world plane intersect at vanishing points in the image, and the vanishing points lie on $\mathbf{l}_\infty$. The vanishing line is the cross product of two vanishing points $\mathbf{l}_\infty = \mathbf{v_1}$ x $\mathbf{v_2}$; it is homogeneous and has two degrees of freedom. Now $\mathbf{H}_p$ is determined as:

$$\mathbf{H}_p = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_{\infty 1} & l_{\infty 2} & l_{\infty 3} \end{pmatrix} \tag{2.5}$$

By applying this matrix, the image can be affine rectified (Figure 5).

In the case of a pair of orthogonal directions, it is possible to further stratify the rectification by decomposing $\mathbf{H}_a$ as:

$$\mathbf{H}_a = \mathbf{A}_2 \mathbf{A}_1 \mathbf{R}_a \tag{2.6}$$

To parameterize $\mathbf{A}_2$, $\mathbf{A}_1$ and $\mathbf{R}_a$, let us consider two orthogonal vanishing points $\mathbf{v_1}$ and $\mathbf{v_2}$. The effect of $\mathbf{H}_p$ on the vanishing points is to transform them to the form $\mathbf{v_{1A}} = (x, x, 0)^T$, where they define directions. $\mathbf{v_{1A}}$ can be written as a unit norm direction vector $\mathbf{v_{1A}} = (\cos(\varphi), \sin(\varphi), 0)^T$, where φ is the angle of $\mathbf{v_{1A}}$ with the horizontal axis. Now $\mathbf{R}_a$ is a rotation matrix that rotates $\mathbf{v_{1A}}$ to the horizontal axis:

$$\mathbf{R_a} = \begin{pmatrix} \cos(\phi) & \sin(\phi) & 0 \\ -\sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.7}$$

If the angle between $\mathbf{v_{1A}}$ and $\mathbf{v_{2A}}$ is θ, $\mathbf{v_{2A}}$ creates an angle of π – θ with the vertical axis and the transformation

$$\mathbf{A}_1 = \begin{pmatrix} 1 & -\cot(\theta) & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.8}$$

transforms a second direction to the vertical axis without changing the orientation of the horizontal axis.

The final transformation $\mathbf{A}_2$ is:

$$\mathbf{A}_2 = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad (2.9)$$

where λ, μ are used to correct the relative scale in horizontal and vertical directions, respectively. The length of the line segment that has been chosen to crop the photo is measured and then, the length of the same segment after the transformation is measured and the scale factor μ is determined. The other scale factor λ, is determined with the same way but now with respect to the horizontal axis.

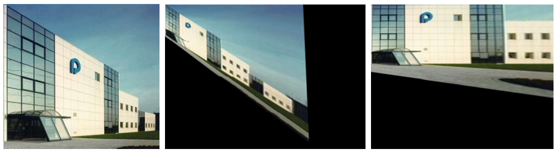

(a) (b) (c)

Figure 5. (a) Source image. (b) Affine rectification, after applying $\mathbf{H}_{p.}$ (c) after applying $\mathbf{H}_a \cdot \mathbf{H}_p$.

2.5 3D Reconstruction

At this final stage, all the necessary information is known so as to reconstruct the image. Points in the rectified images are being matched to the 3D coordinates that have been obtained after applying the matrix **N**. Figure 6 shows the 3D reconstruction of Figure 2 and Figure 3. Figure 6 (c) shows another example of a reconstructed building.

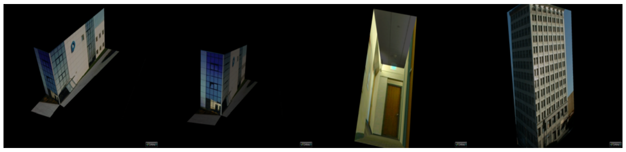

(a) (b) (c)

Figure 6. (a) 3D reconstruction of Figure 2. (b) 3D reconstruction of Figure 3. (c) 3D of a building.

3. CONCLUSIONS

In this paper a novel algorithm is proposed, able to create a 3D reconstruction of a scene, fully automated, with no user interaction. Two are the main advantages of this work: firstly, the proposed algorithm can be used for accurate real-time 3D reconstructions of both interior and exterior buildings; secondly, it is fully automated and no user interaction is needed in any step of the algorithm. Further research is needed in order to be able to recognize more objects or humans inside a scene so as to fully reconstruct a scene using only one image.

4. ACKNOWLEDGMENTS

This work was supported by the EU FP7 project 3DLife Network of Excellence project, ICT-247688. The algorithm for the vanishing point detection was developed by the University of Valladolid, Spain.

5. REFERENCES

[1] J. M. Coughlan and A. L. Yuille. Manhattan world: Compass direction from a single image by Bayesian inference. In *ICCV '99: Proceedings of the International Conference on Computer Vision- Volume 2, page 941, Washington, DC, USA, 1999.* IEEE Computer Society. DOI= http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.33.5078 .

[2] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. *Int. J. Computer Vision,* 40(2):123-148, 2000. DOI= http://portal.acm.org/citation.cfm?id=365888 .

[3] A. Criminisi, I. Reid, and A. Zisserman. A plane measuring device. *Image and Vision Computing,* 17(8):625-634, 1999.

[4] E. Delage, H. Lee, and A. Y. Ng. A dynamic Bayesian network model for autonomous 3d reconstruction from a single indoor image. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on,* 2:2418-2428, 2006. DOI= http://www.stanford.edu/~hllee/cvpr06_3dReconIndoorScene.pdf

[5] Google Sketch Up. http://sketchup.google.com/

[6] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, 2003.

[7] D. Hoiem, A. Efors, and M. Hebert. Geometric context from a single image. In *Proceedings of IEEE Conference Computer Vision and Pattern Recognition,* 2005. DOI = http://www.cs.uiuc.edu/homes/dhoiem/publications/Hoiem_Geometric.pdf

[8] J. Huang and B. Cowan. Simple 3D Reconstruction of Single Indoor Image with Perspective Cues. crv, pp.140-147, 2009 *Canadian Conference on Computer and Robot Vision.* DOI= http://www.computer.org/portal/web/csdl/doi/10.1109/CRV.2009.33

[9] D. C. Lee, M. Hebert and T. Kanade. Geometric Reasoning for Single Image Structure Recovery. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, June, 2009. DOI= www.cs.cmu.edu/~dclee/pub/cvpr09lee.pdf

[10] D. Liebowitz, Antonio Criminisi, and Andrew Zisserman. 1999. Creating Architectural Models from Images. *In Proc. EuroGraphics, vol.18, 1999*. DOI= http://www.robots.ox.ac.uk/~vgg/publications/papers/liebowitz99.pdf .

[11] D. Liebowitz, *Camera Calibration and Reconstruction of Geometry from Images*. Merton College Robotics Research Group Department of Engineering Science University of Oxford Trinity Term 2001. DOI= http://www.robots.ox.ac.uk/~vgg/publications/papers/liebowitz01.pdf

[12] C. Rother. A new approach for vanishing point detection in architectural environments. In *BMVC,* pages 382-391, 2000.

DOI= www.bmva.org/bmvc/2000/papers/p39.pdf

[13] A. Saxena, S. H. Chung, and A. Y. Ng. Learning depth from single monocular images. In *Neural Information Processing Systems (NIPS),* 2005. DOI = http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.72.8799