An Implementation of Single-View Metrology

Shih-Yu Sun
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
sysun@mit.edu

Abstract

The single-view metrology method extracts three-dimensional information from a single image based on projective geometry. As this method uses vanishing points found in the scene, it is especially suitable for scenes that involve man-made structure, like architectures and geometric elements. In this paper, an implementation of the single-view metrology method is described. The two main components of the method, planar homography and measurement of distances from the reference plane, were developed and tested using images with real scenes. Also, algorithms were implemented to increase the level of automation for vanishing point detection. Finally, examples of complete three-dimensional reconstruction of single views are demonstrated.

1. Introduction

The single-view metrology method makes use of planes and parallel lines in the scene to extract the physical dimensions of structures from a single perspective view, given minimal prior knowledge about the scene. The prior knowledge usually involves the recognition of vanishing points of a reference plane, and a vanishing point of a direction not parallel to the reference plane. As a result, this method is especially suitable for scene structures that involve parallel lines, which could often be found in man-made structures, such as architectures and geometric elements. Concepts of the single-view metrology method were described in a number of papers [1]-[3] and were later generalized by A. Criminisi et al. [4].

In this paper, an implementation of three-dimensional (3D) model reconstruction based on the single-view metrology method is described. A reference plane is chosen in the image, which, if applicable, is usually the ground floor. The elements in the scene are first segmented to provide the texture maps. Finally, element positions on the reference plane and distances from that plane (i.e. heights, if the ground plane is chosen as the reference) are estimated. This estimation mostly follows the single-view metrology algorithms suggested in [5], and could be performed in an uncalibrated framework, in which no estimation of camera pose or the internal parameters is required.

The remaining of this paper is organized as follows. In Section 2 and Section 3, the approaches for estimating object positions and heights are described, respectively, and the experimental results are shown. In Section 4, examples of complete 3D reconstruction are demonstrated.

2. Positions on the reference plane

For a projective camera, the world coordinates of a plane in the scene and tbhe corresponding image coordinates are related by a 3-by-3 homography matrix. [6] This relationship can be expressed by the following equation in homogeneous coordinates:

$$\mathbf{X} = \mathrm{H}\mathbf{x} \tag{1}$$

where $\mathbf{X}$ is a point on the world plane, $\mathbf{x}$ is the corresponding point in the image plane, and H is the image-to-world homography matrix.

Since H has eight degrees of freedom in the homogeneous representation, it can be determined by at least four corresponding points. Writing H as a 9-vector $\mathbf{h} = [h_{11}\, h_{12}\, h_{13}\, h_{21}\, h_{22}\, h_{23}\, h_{31}\, h_{32}\, h_{33}]^T$, the constraints given by *n* corresponding points, $\mathbf{X_i}$ and $\mathbf{x_i}$ for i = 1, 2, 3, ……*n*, can be written in a compact form: $\mathrm{A}\mathbf{h} = 0$, where A is a $2n \times 9$ matrix:

$$\mathrm{A} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{y}_1 & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{x}_1\mathbf{X}_1 & -\mathbf{y}_1\mathbf{X}_1 & -\mathbf{X}_1 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{x}_1 & \mathbf{y}_1 & \mathbf{1} & -\mathbf{x}_1\mathbf{Y}_1 & -\mathbf{y}_1\mathbf{Y}_1 & -\mathbf{Y}_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{x}_n & \mathbf{y}_n & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{x}_n\mathbf{X}_n & -\mathbf{y}_n\mathbf{X}_n & -\mathbf{X}_n \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{x}_n & \mathbf{y}_n & \mathbf{1} & -\mathbf{x}_n\mathbf{Y}_n & -\mathbf{y}_n\mathbf{Y}_n & -\mathbf{Y}_n \end{bmatrix} \tag{2}$$

Up to a scaling factor, the solution $\mathbf{h}$ is the unit eigenvector of the matrix $\mathrm{A}^T\mathrm{A}$ with the minimum eigenvalue. Derivation of the solution can be found in [5].

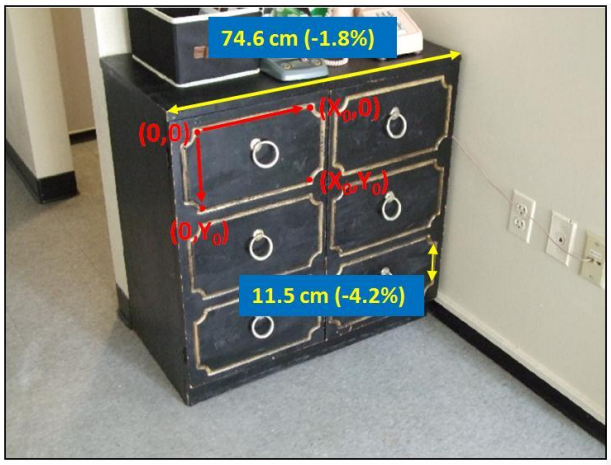


Figure 1: Four corners of a rectangle were used to establish the world coordinate system of the plane (red), and determine the image-to-world homography matrix. Lengths of the segments (yellow) were computed using the homography matrix, where the percentage of error is shown in the parenthesis.

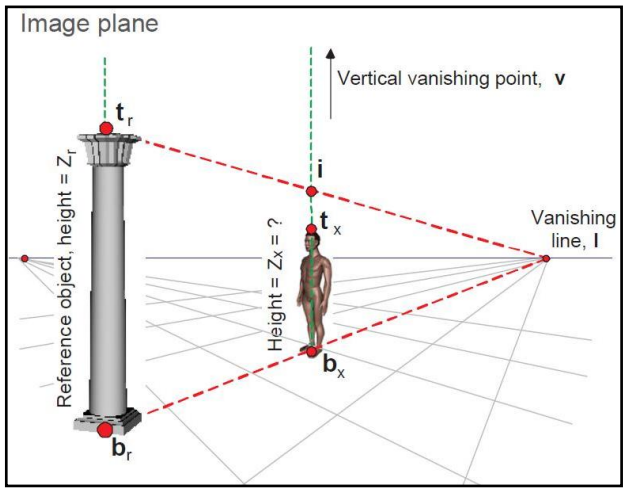


Figure 2: Illustration of height measurement using vanishing points and a reference height. (adapted from [5])

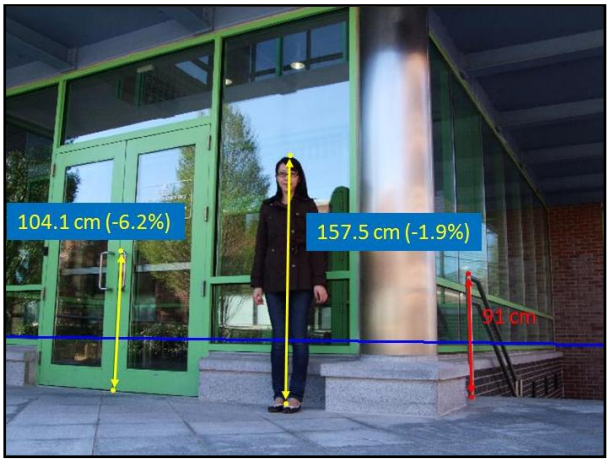


Figure 3: The blue line shows the horizontal vanishing line, which was found from the parallel lines in the scene. The stick marked in red was measured as the reference height. The heights of the girl and the right door handle were estimated (yellow), where the percentage of error is shown in the parenthesis.

2.1. Experimental results

Given the image-to-world homography matrix H, the world coordinates of points on the world plane and the physical dimensions can be computed. The image shown in Fig.1 was used for verifying the performance of this planar homography method. The frontal side of the cabinet is approximately planar. The four corners of a rectangle were selected as the corresponding points (marked in red), and the sides of the rectangle, X_0 and Y_0, were measured to establish the world coordinate system.

The homography matrix H found from the correspondences allows the estimation of physical dimensions on the plane. For example, in Fig.1, the lengths of the two line segments marked in yellow were estimated by the homography matrix. It can be seen that the estimates are fairly accurate.

3. Distance from the reference plane

The heights of scene elements could be estimated based on vanishing points and a known reference height found in the scene. See Fig.2 for an illustration of the method. Here **l** is the horizontal vanishing line, **v** is the vertical vanishing point, and Z_r is the reference height. $\mathbf{b_r}$ and $\mathbf{t_r}$ are the bottom and top of the reference height, respectively. The metric factor α can then be found by the following equation:

$$\alpha Z_r = -\frac{\|\mathbf{b_r} \times \mathbf{t_r}\|}{(\mathbf{l} \cdot \mathbf{b_r})\|\mathbf{v} \times \mathbf{t_r}\|} \qquad (3)$$

For a given object on the reference plane, its height Z_x can then be found by the following equation, where $\mathbf{b_x}$ and $\mathbf{t_x}$ are the bottom and the top of the object, respectively:

$$Z_x = -\frac{\|\mathbf{b_x} \times \mathbf{t_x}\|}{\alpha(\mathbf{l} \cdot \mathbf{b_x})\|\mathbf{v} \times \mathbf{t_x}\|} \qquad (4)$$

Proof of equation (3) and (4) can be found in [7].

3.1. Experimental results

The measurement of heights by using vanishing points

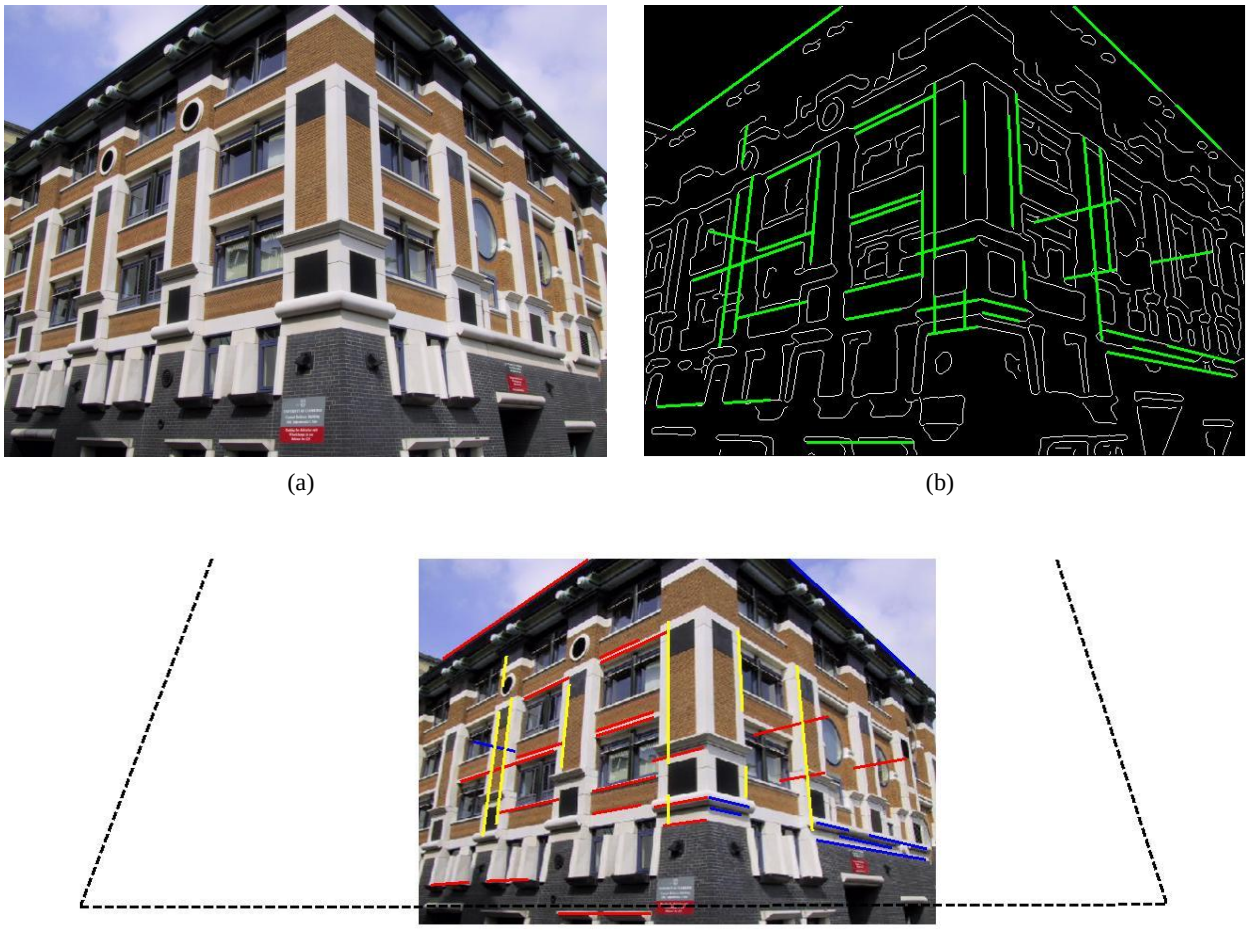


Figure 4: Example of successful automatic detection of vanishing points. The results of the Canny edge detector and the Hough transform applied on the image in (a) are shown in (b), where the line segments found in the Hough transform are marked in green. In (c), groups of line segments found in RANSAC are marked in different colors. The three vanishing points found are linked by dashed black lines. (The vertical vanishing point is not shown.)

was tested on the picture shown in Fig.3, in which the horizontal and vertical vanishing points were found by the sets of parallel lines that can be found in the building. The blue line shown in Fig.3 is the horizontal vanishing line. The height of the metal stick beside the pillar was measured as the reference height. Here the parallel lines and the reference height, as well as the objects whose heights were to be estimated, were manually defined.

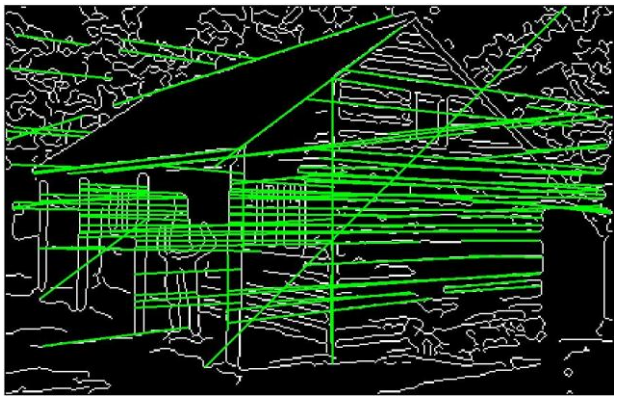
Here the heights of the girl and the right door handle were estimated using the vanishing points and the reference height, as shown in Fig.3, in which the percentage of error is shown in the parenthesis. The estimates are reasonably accurate for the purpose of 3D reconstruction.

3.2. Automatic detection of vanishing points

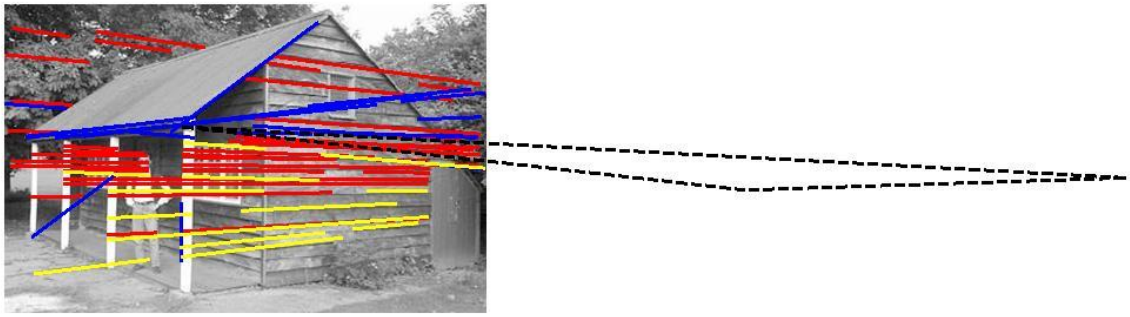
The algorithm suggested in [5] for automatic detection of vanishing points was implemented in an attempt to accelerate the 3D reconstruction process. The Canny edge detector and the Hough transform are first applied on the image to detect straight line segments. Grouping of the lines is then performed based on RANSAC, in order to find sets of lines that intersect at the same points in the image plane. Ideally, for a scene that contains a rich collection of parallel lines along the three axes of the world coordinate system, the top three groups with the most line segments would give the vanishing points for those axes.

It was found, however, that this simple algorithm was not very robust. A number of parameters that are involved

(a)



(b)



(c)

Figure 5: Example of failed automatic detection of vanishing points. The results of the Canny edge detector and the Hough transform applied on the image in (a) are shown in (b), where the line segments found in the Hough transform are marked in green. In (c), groups of line segments found in RANSAC are marked in different colors. The three vanishing points found are linked by dashed black lines.

in the Canny edge detector and the Hough transform need to be carefully fine-tuned for a given image to successfully find the vanishing points. Fig.4 and Fig.5 show successful and failed examples of this algorithm, respectively. In the failed example (Fig.5), it can be seen that the Hough transform incorrectly classified some groups of pixels as straight line segments, and an intersection point of many line segments (in blue) was incorrectly considered a vanishing point. Moreover, it appeared that the groups of lines marked in red and yellow should give the same vanishing point, but two separate points were found from these lines.

Given the lack of robustness of this simple algorithm for vanishing point detection, it was determined that the parallel lines in the scenes would be defined manually using a graphical user interface with zoom-in functionality.

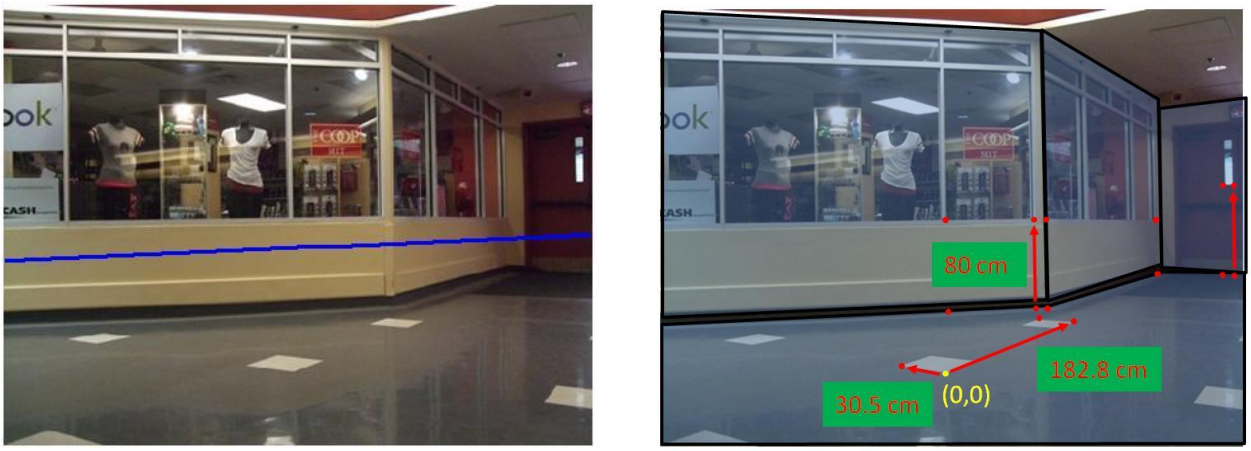
4. Complete 3D Reconstruction

Using the methods for estimating the positions and heights of scene elements in single-view scenes, complete 3D reconstruction could be performed. The ground plane is often selected as the reference plane. Scene elements on the reference plane, such as people and architectural structures, are manually segmented. The positions and heights of the elements are estimated, and the segmented portions of the image are used as the texture maps for the 3D model.

4.1. Example: MIT COOP in the student center

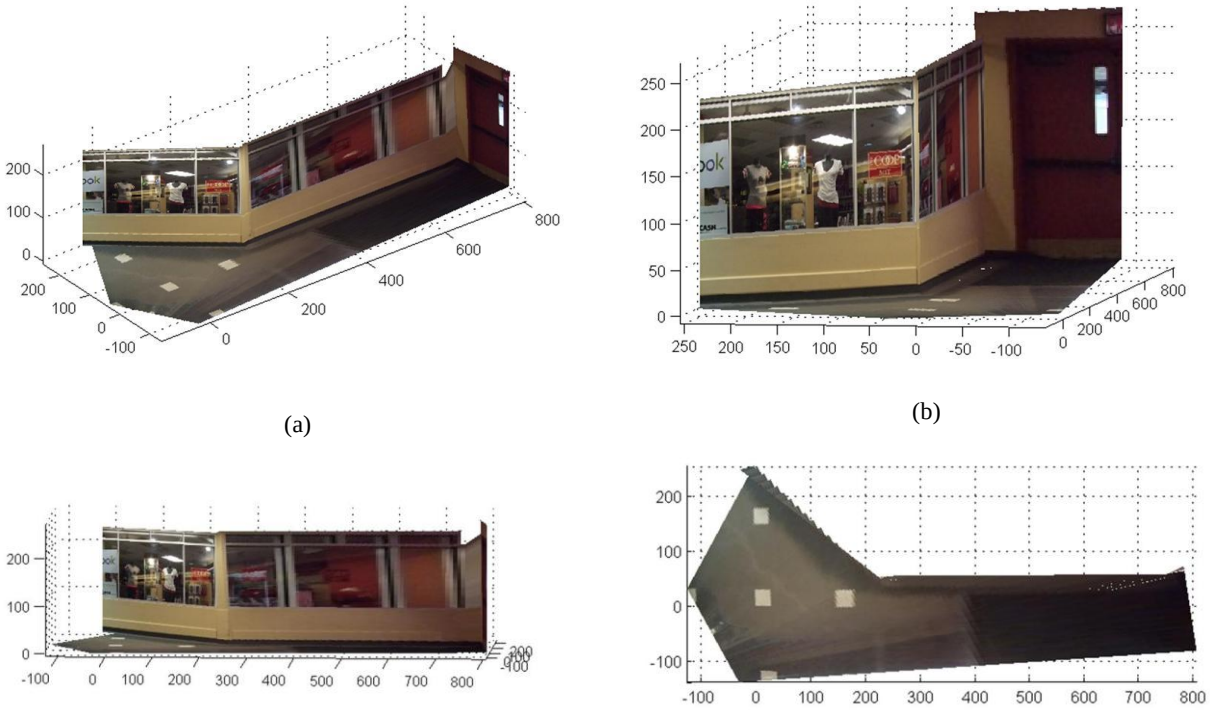
A picture taken in front of the MIT COOP in the Stratton student center was used to provide the scene for a complete 3D reconstruction, as shown in Fig.6(a). The parallel lines in the scene were manually defined for finding the vanishing points.

Fig.6(b) shows the required manual input after the vanishing points are found. The ground plane was chosen as the reference plane, and two orthogonal line segments (marked in red) were measured to establish the world coordinate system. In the frontal plane of the COOP, the height of the window was measured as the reference height for estimating the planar homography matrices for the two



(a) (b)

Figure 6: Complete 3D reconstruction was performed for the scene in front of the MIT COOP (a). The blue line is the horizontal vanishing line. (b) illustrates the manual input for 3D reconstruction after the vanishing points were estimated, including distance measurement, point selection for homography estimation, and segmentation. The selected origin for the world coordinate system is marked in yellow.



(a) (b)

(c) (d)

Figure 7: The 3D model of the scene shown in Fig.6 in different viewpoints. (unit: cm)

planes of the COOP. This reference height was also used to obtain an estimate of the height of the door window at the right of the scene (marked in red), which was 135.8 cm with a +5.4% error. Using these measurements, four corresponding points were selected to estimate the homography matrix for each segmented plane in the scene. Note that the dimensions of the scene were fully determined by these measurements, so the dimensions of a

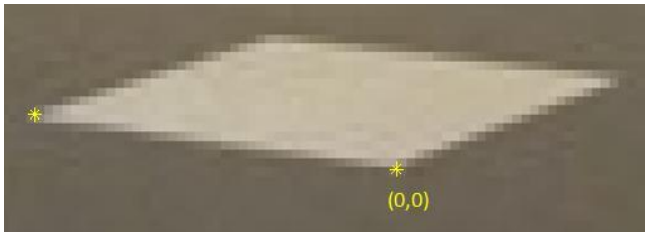


Figure 8: Fig.6(b) zoomed in to the portion that contains the origin of the world coordinate system.

given portion of the scene can be directly read off from the reconstructed 3D model.

The reconstructed 3D model is shown in Fig.7 in different viewpoints, where some artifacts can be seen. In Fig.7(b), the door appears to be distorted. In Fig.7(c), the windows also appear to be distorted, and there are noticeable artifacts between the window plane and the door plane. These artifacts should be mainly due to the limited precision of point selection in the image and the resulting errors in the estimation of homography matrices. Fig.8 shows a zoom-in view of Fig.6(b), where it can be seen that the corners of the rectangle are not perfectly defined in the image due to the limited spatial resolution. As a result, errors from defining corresponding points were inevitable.

4.2. Example: MIT 150 FAST Light

The single-view 3D reconstruction method can be applied to scenes where there are patterns that are barely recognizable from the available viewpoints. For example, Fig.9(a) shows words that were projected onto the ground, which was part of the MIT 150 FAST Light activities. Given the highly tilted viewing angle, the words could barely be recognized. However, thanks to the square bricks on the ground, the homography matrix for the ground plane could be estimated for a 3D reconstruction of the ground. See Fig.9(b) for a view of the ground from the top, which clearly shows "astounding alumni."

It should be noted that in this case, no measurement was made from the scene, and it was assumed that the side of the square bricks is one unit long when establishing the world coordinate system. Therefore, unlike the MIT COOP example shown previously, this 3D model was measured in the unit of "brick side", and the real dimensions are not available.

5. Conclusions

This paper demonstrates examples of planar homography and measurement of object heights using vanishing points. The estimates were shown to be fairly accurate, given that point and line selection in the image was performed manually. Attempts were made to automate

(a)

(b)

Figure 9: The 3D reconstruction of the ground plane in the MIT FAST Light example, where the tilted viewing angle in (a) is removed to give the top view in (b).

vanishing point detection using the Canny edge detector, the Hough transform, and RANSAC-based line grouping, but the algorithm was found to lack in robustness.

An example of complete 3D reconstruction was shown for the scene in front of the MIT COOP. The artifacts and possible sources of errors were discussed. The application of 3D reconstruction in the MIT FAST Light activities was also demonstrated. In addition to the above applications, the single-view metrology method could also be useful in architecture, surveillance, and forensic sciences.

References

[1] I. Reid and A. Zisserman, "Goal-directed video metrology," *Computer Vision—ECCV'96*, 1996, p. 647–658.

[2] T. Kim, Y. Seo, and K.S. Hong, "Physics-based 3D position analysis of a soccer ball from monocular image sequences,"

Computer Vision, 1998. Sixth International Conference on, 1998, p. 721–726.

[3] Y. Horry, K.I. Anjyo, and K. Arai, “Tour into the picture: using a spidery mesh interface to make animation from a single image,” *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, 1997, p. 225–232.

[4] A. Criminisi, I. Reid, and A. Zisserman, “Single view metrology,” *International Journal of Computer Vision*, vol. 40, 2000, p. 123–148.

[5] A. Criminisi, “Single-view metrology: Algorithms and applications,” *Lecture notes in computer science*, 2002, p. 224–239.

[6] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge university press, 2008.

[7] A. Criminisi, *Accurate visual metrology from single and multiple uncalibrated images*, Springer Verlag, 2001.