Proceedings of the IIEEJ Image Electronics
and Visual Computing Workshop 2012
Kuching, Malaysia, November 21-24, 2012

THREE-DIMENSIONAL RECOVERY OF BUILDINGS ENVIRONMENT UNDER MANHATTAN-WORLD CONSTRAINT

Yasuhiro Ohyama† *Joo Kooi Tan*† *Hyoungseop Kim*† *Seiji Ishikawa*†

†Department of Control Engineering, Kyushu Institute of Technology
Tobata, Kitakyushu 804-8550, Japan.

ABSTRACT

This paper proposes a 3-D recovery method of an environment containing buildings. The difficulties of buildings recovery include little texture regions and a large scale nature, resulting in the accumulation of recovery errors. The proposed method introduces the Manhattan-world constraint in the algorithm of structure from motion (SfM) which is employed for the shape recovery from a set of sparse feature points. Experimental results show that the proposed method achieves higher accuracy in the recovery compared to the existent methods based only on the SfM algorithm.

1. INTRODUCTION

In recent years, 3D TVs, augmented reality, video games and various other technological fields that employ 3D-shape recovery have become much more popular than ever. Among them, buildings and other static objects recovery has been studied enthusiastically as one of the important research subjects in the computer vision community. One of the main difficulties in the recovery of such objects is that the accuracy of the reconstruction often drops under textureless regions or large scale objects.

There are three main methods of buildings environment recovery. Sparse recovery [2] normally employs the Structure from Motion (SfM) techniques. But it is weak to positional noise, since the recovery depends on sparsely distributed feature points. Model-based recovery [3] tries to fit and modify box, e.g., models to buildings. Dense recovery [4] makes a depth image which gives depth information of all the pixels in an image employing the result of the sparse recovery. The Manhattan-world constraint [1], defined in the next section, is normally employed in the model-based recovery and the dense recovery to achieve precise shape recovery and not employed in the sparse recovery. However, since the latter two recovery techniques often use the result of the sparse recovery of the object concerned, it is important to realize precise recovery even by the sparse recovery. The idea of the present paper is therefore the employment of the Manhattan-world constraint into the sparse recovery.

In the present paper, we propose a method of recovering a static environment containing buildings with high accuracy by the employment of the Manhattan-world constraint into a SfM algorithm.

2. MANHATTAN-WORLD CONSTRAINT

The Manhattan-world constraint is defined as follows;
A1: All the planes in the Manhattan-world are perpendicular to one of the three coordinate axes by which the Manhattan-world is described.
A2: Every point in the Manhattan-world lies on one of the above planes.

3. PROPOSED METHOD

The outline of the proposed method is shown in **Fig. 1**. The procedure is composed of two stages. In the first stage, an initial model of the environment containing buildings is created by a SfM algorithm constrained by the Manhattan-world. In the second stage, the initial model is extended by concatenation so that it may match newly fed image data of the environment.

3.1 Recovery of the locations of 3-D points

Given a pair of successive image frames I_t and I_{t+1}, feature points are extracted on image I_t using the GLOH (Gradient Location and Orientation Histogram) [5] which is invariant to scaling, rotation and illumination change. The feature points are then tracked on image I_{t+1} by the L-K tracker [6] to find correspondence.

Employing pairs of corresponding feature points between I_t and I_{t+1}, a camera rotation matrix R and a parallel translation vector $\boldsymbol{T} \equiv (T_x \; T_y \; T_z)^{\mathrm{T}}$ are computed using the

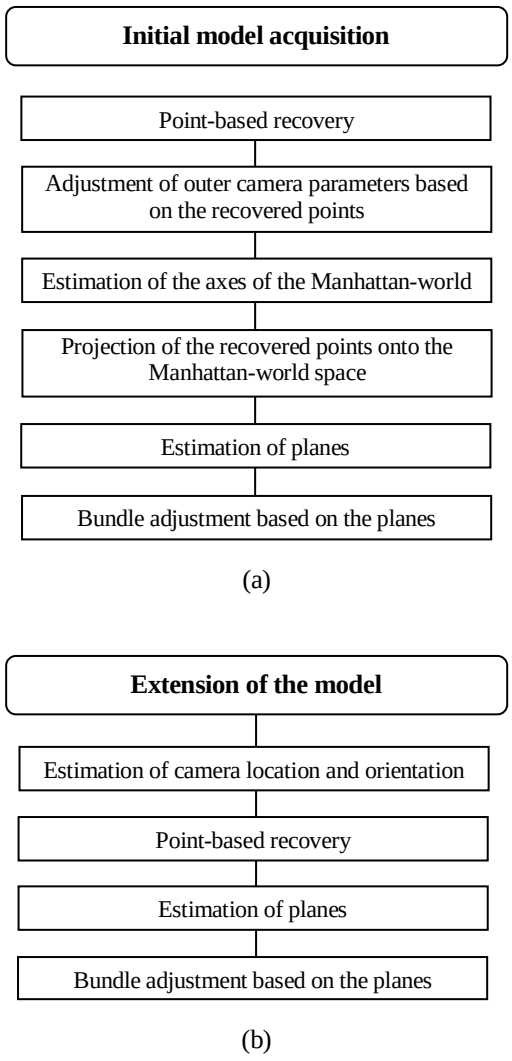

Fig. 1 Outline of the proposed method: (a) Acquisition of an initial model, (b) extention of the model.

epipolar geometry. Namely, the fundamental matrix F is computed using the set of corresponding feature points and then the essential matrix E is obtained from the relation $E=A^{\mathrm{T}}FA$, where A is a camera inner parameter matrix computed in advance. By the decomposition of the matrix E, we have R and $\boldsymbol{T}$ [7]. Then, if we denote the projected point of a point $\boldsymbol{X}_i$ in the 3-D space on image I_t by $\boldsymbol{m}_{t,i}$ and on image I_{t+1} by $\boldsymbol{m}_{t+1,i}$, their relations are formulated as

$$\begin{aligned} \lambda \boldsymbol{m}_{t,i} &= A(I\ \boldsymbol{0})\boldsymbol{X}_i \\ \lambda \boldsymbol{m}_{t+1,i} &= A(R\ \boldsymbol{T})\boldsymbol{X}_i \end{aligned} . \qquad (1)$$

By solving Eq.(1), the point $\boldsymbol{X}_i$ recovers its 3-D location. Here λ is a real constant.

Refinement of the rotation matrix R and the parallel translation vector $\boldsymbol{T}$ is performed employing the recovered 3-D points $\{\boldsymbol{X}_i\}$ and their projection $\{\boldsymbol{m}_{t+1,i}\}$ on image I_{t+1}. The DLT method is employed in the first place and then nonlinear optimization aiming at minimization of re-projection errors is performed by Levenberg-Marquard method for further refinement.

3.2 Deriving the Manhattan-world coordinate system

To obtain a 3-D structure, the Manhattan-world, from an image, line segments are extracted in the first place. A useful edge detector [8] is employed to get an edge image and line segments are obtained by connecting those edge segments whose gradients are mutually close on a line. **Figure 2a** shows an example of the line segments detected from an edge image.

Since the Manhattan-world contains horizontal planes and the planes perpendicular to them, vanishing points of the detected line segments define the axes of an orthogonal coordinate system. The intersections, denoted by $\boldsymbol{v}_j$ (j=1,2,…,J), of all the pairs of the detected line segments are calculated and they are merged into several groups by evaluating the likelihood defined by

$$D(\boldsymbol{v}_j, \varepsilon_i) = dist(\boldsymbol{e}_i^1, [\bar{\boldsymbol{e}}_i]_{\mathrm{X}} \boldsymbol{v}_j) . \qquad (2)$$

This is the likelihood that a line segment ε_i passes through the vanishing point $\boldsymbol{v}_j$. It is defined by the distance between the end point $\boldsymbol{e}_i^1$ which is far from $\boldsymbol{v}_j$ on the line segment ε_i and the line connecting $\boldsymbol{v}_j$ and the middle point $\bar{\boldsymbol{e}}_i$ of ε_i. Vanishing point $\boldsymbol{v}_k$ is included in a set of vanishing points $\{\boldsymbol{v}_j\}$, if they have common line segments whose values of D are smaller than a specified threshold. By repeating this merging process, the vanishing points are finally classified into a small number of groups. The line segments shown in Fig. 2a are collected into several groups as depicted in Fig. 2b, where identical color line segments make a group.

Now we have to choose three orthogonal axes defining the Manhattan-world. A vanishing point $\boldsymbol{v}_j$ on an image is inversely projected to a 3-D point $A^{-1}\boldsymbol{v}_j$ in the camera coordinate system. For i, j (i, j =X,Y,Z; $i \neq j$), the condition of the orthogonality is written as

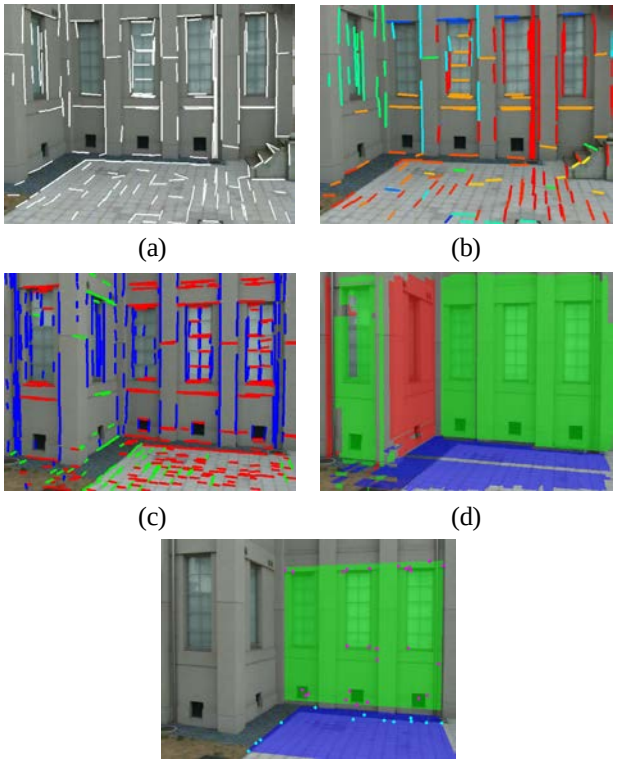

Fig. 2 Deriving the Manhattan-world: (a) Extracting line segments, (b) the line segments classified w.r.t. several vanishing points, (c) three classes of the line segments passing through the chosen three vanishing points which provide the coordinates of the Manhattan-world, (d) the regions corresponding to the three vanishing points, and (e) recovered planes in the Manhattan-world.

$$(A^{-1}\boldsymbol{v}_i,\ A^{-1}\boldsymbol{v}_j)=0. \tag{3}$$

Since this equation does not hold exactly on account of the positional noise included in $\boldsymbol{v}_j$, three vanishing points which make the left-hand side of Eq.(3) the minimum are chosen as $\boldsymbol{v}_X$, $\boldsymbol{v}_Y$ and $\boldsymbol{v}_Z$. Then vectors

$$\boldsymbol{V}_X \equiv A^{-1}\boldsymbol{v}_X, \boldsymbol{V}_Y \equiv A^{-1}\boldsymbol{v}_Y, \boldsymbol{V}_Z \equiv A^{-1}\boldsymbol{v}_Z \tag{4}$$

are the coordinate axes of the Manhattan-world defined from the image concerned. According to the chosen three vanishing points, the line segments shown in Fig. 2b are classified into three as illustrated in Fig. 2c.

By the employment of the axes $\boldsymbol{V}_X$, $\boldsymbol{V}_Y$ and $\boldsymbol{V}_Z$, the coordinate system of the Manhattan-world is defined by the matrix M_{M} as follows;

$$M_{\mathrm{M}} = \begin{pmatrix} \boldsymbol{V}_X & \boldsymbol{V}_Y & \boldsymbol{V}_Z & \boldsymbol{0} \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{5}$$

The recovered 3-D points and the rotation matrix R and the parallel translation vector $\boldsymbol{T}$ of the observing camera are all transformed into those in the Manhattan-world coordinate system by

$$M_{\mathrm{M}} = P_{\mathrm{M}} M_{\mathrm{W}} \tag{6}$$

where P_{M} is a transformation matrix from the world coordinates M_{W} to the Manhattan-world coordinates M_{M}.

3.3 Recovery of planes

Regions representing the planes perpendicular to respective coordinate axes of the Manhattan-world are extracted on image I_t. The idea is shown in Fig. 2d where the region which does not include the line segments toward the vanishing point $\boldsymbol{v}_X$, indicated by red line segments in Fig. 2c is extracted (indicated by red). In the same way, the region which does not include the line segments toward the vanishing point $\boldsymbol{v}_Y$, indicated by green line segments in Fig. 2c is extracted (indicated by green) and the region which does not include the line segments toward the vanishing point $\boldsymbol{v}_Z$, indicated by blue line segments in Fig. 2c is extracted (indicated by blue). This image is called an orientation map.

The feature points chosen initially on image I_t are categorized on the orientation map to one of those regions. With each region, the 3-D points corresponding to the feature points categorized in the region are examined their locations toward the axis perpendicular to the plane which the region represents. If the number of the 3-D points exceeds a given threshold at a certain position on the axis, a plane is computed from the points. Examples of the recovered planes are shown in Fig. 2e.

3.4 Extension of the model

The Manhattan-world model created by the above procedure is an initial model. This model is extended each time the camera provides a new key frame.

Once a new key frame is fed, the location and the orientation of the camera are computed by solving the perspective n-point problem. Then newly observed feature points recover their 3-D locations from which planes are computed using the branch-and-bound algorithm [8]. Finally the planes, and camera location and orientation are adjusted again based on the minimization of the re-projected errors by the employment of a non-linear optimization method.

4. EXPERIMENTAL RESULTS

Employing the proposed method, experiments were conducted in a real-world environment. Part of a building was taken a video by a handheld camera. The key image frames employed for the recovery is 13 out of 317 frames: Recovered feature points are 560: Recovered planes are 5: The computation time is 37.8 second by a PC with a 3.20 GHz CPU. **Figure 3** shows part of the results: (a) The input images from four viewpoints, (b) results of the projection of the recovered points and planes into the 3-D world, and (c) the obtained depth maps which show the distance of every pixel in the image from the observing camera. As shown in **Table 1**, comparison of the proposed method (SfM employing the Manhattan-world constraint) with the conventional method (only SfM) was done with respect to the rate of outliers existent in the recovered points. It was shown that the proposed method outperformed the existent method in the two outdoor cases.

5. DISCUSSION AND CONCLUSIONS

A method was proposed for recovering 3-D shape of a buildings environment by a SfM technique with the Manhattan-world constraint which was not done before.

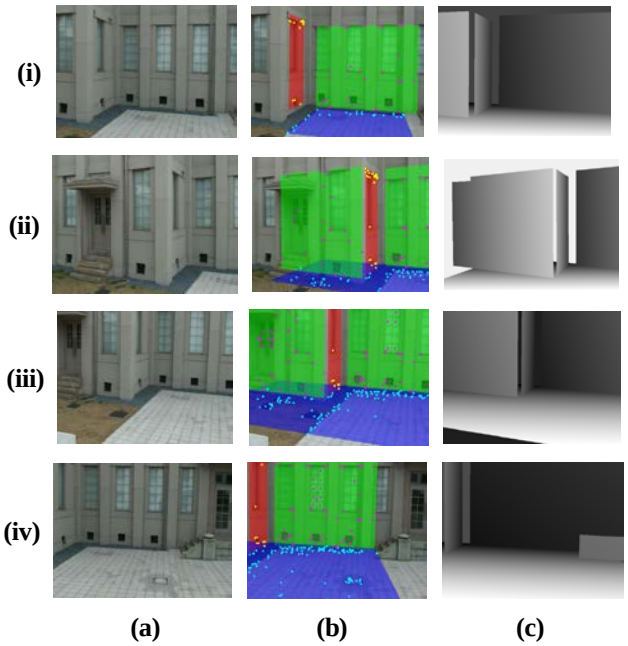

Fig. 3 Experimental Results: (a) Input images, (b) result of the projection of the recovered points and planes, (c) depth maps estimated from the plane: (i) Viewpoint 1, (ii) viewpoint 2, (iii) viewpoint 3, (iv) viewpoint 4.

Table 1 Performance of the proposed method.

Experiment	Outdoors_1		Outdoors_2	
Method	SfM	SfM+MW	SfM	SfM+MW
Rate of outliers [%]	2.60	0.89	0.89	0.39

SfM: Structure from motion, MW : Manhattan-world.

Table 1 shows that the proposed method reduces the rate of outliers better than the conventional method. This may be because the proposed method discards the recovered erroneous points by the adjustment of the 3-D points based on the estimation of the planes. Therefore less number of outliers can improve the solution of the PnP problem by realizing more reliable estimation of the camera position and rotation, resulting in more rigorous 3-D recovery of a buildings environment.

The proposed method is for sparse recovery. The recovery with higher precision on this stage may be advantageous for dense recovery employing the result of the sparse recovery.

REFERENCES

[1] J. M. Coughlan, A. L. Yuille, "Manhattan world: Orientation and outlier detection by bayesian inference", *Neural Computation*, pp. 1063-1088, 2003.

[2] M. Pollefeys, et al., "Visual modeling with a hand-held camera", *International Journal of Computer Vision*, pp. 207-232, 2004.

[3] C. A. Vanegas, et al., "Building reconstruction using manhattan-world grammars", *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 358-365, 2010.

[4] Y. Furukawa, et al., "Manhattan-world stereo", *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1422-1429, 2009.

[5] K. Mikolajczyk, C. Schmid, "A performance evaluation of local descripters", *IEEE Trans. on PAMI*, vol.27, no.10, pp. 1615-1630, 2005.

[6] B. Lucas, T. Kanade, "An iterative image registration technique with an application to stereo vision", *Proc. Int. J. Conf. on Artificial Intelligence*, pp. 121-130, 1981.

[7] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge Univ. Press, 2004.

[8] D.C. Lee, M. Hebert, T. Kanade: "Geometric reasoning for single image structure recovery", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2009.