Semi-metric reconstruction from a single image using orthogonality and parallelism

Jun-Sik Kim† and In So Kweon ‡

Department of Electrical Engineering and Computer Science, KAIST, Daejeon, Korea
†jskim@rcv.kaist.ac.kr, ‡iskweon@kaist.ac.kr

Abstract In real world, there are many artificial objects. We propose a new method to reconstruct the 3D structure of the artificial objects from the scene pictures using orthogonality and parallelism. The new transformation group "semi-metric space" is defined and using that we describe the artificial scenes effectively. Through the analysis of the plane and parallax, the relative distance between a line and the reference plane is estimated, and it gives us a partial 3D structure from a single image. The algorithms are verified with real image captured with camera in commercial mobile phone.

1 Introduction

In recent years, imaging systems have come into wide use in public. Most popular imaging systems are digital cameras. Under one thousand dollars, we can get a public digital camera which has four million pixels or more. There are even mobile phones which have imaging capturing modules with about three million pixels.

Even if you do not have a high quality digital camera, you can use high-quality digital images through broadcasts. Images of high-definition television (HDTV) have very high resolution (1920 × 1080), and it is relatively same with one from cameras with two-million effective pixels or more. Cost for capturing HDTV images is under two hundred dollars. In other words, there are so many ways for people to have high quality images in these days with very low cost. Although imaging capturing devices flood into the public, computer vision technologies are rarely distributed. Except making panoramas with images from rotating cameras, people have not used the information of their images using vision algorithms or image analysis. Especially, high-quality images have much geometric information, but extracting the information requires some specific knowledge about projective geometry.

To make geometric information in captured scenes used in public, efforts to simplify the process by users are needed. Especially, manual setting of parameters of the algorithm is critical to simplify the user intervention. We want to make a simple framework to be used easily by the public.

Unfortunately there are few cases to obtain the suggested independent information of the scene. People can detect a parallelism of line set, and furthermore, an orthogonality very easily. In fact, most of visual illusions are based on the properties of the human visual system [1]. It means that human visual systems have been well-trained to detect the parallelism and the orthogonality. But it is quite difficult to find an exact aspect ratio of rectangles in three dimensional space. As pointed out in [2], this kind of information is critical to reveal the correct structure of the captured scenes.

This can be a problem if we want to work with *unknown* scenes. Sometimes, we cannot measure the scene physically and we have no idea about the cameras which are used to capture the scene. Practically, there are many cases that we have some rectangles whose aspect ratios are unknown. For example, these cases occur when we want to use some snapshot images captured in a travel, or captured images from TV signal. What we can use is just information from the human visual system, those are the parallelism and the orthogonality.

In this paper, we study the possibility to use only the parallelism and the orthogonality with one or a few images in reconstructing artificial objects. First, we propose a new transformation space called *semi-metric space*. Next, we investigate a possibility to reconstruct structures of a scene from the captured images using the properties of the semi-metric space. In practice, we extract some partial information of a scene from a single image using the proposed framework with the image captured with a camera in a mobile phone.

2 Semi-metric space

A semi-metric space is a space that is represented by a *semi-metric transformation*. A semi-metric transformation in $\mathbb{P}^2$ is expressed as

$$\mathbf{x}' = \mathsf{H}_{\mathsf{SM}}\mathbf{x} = \begin{bmatrix} s_1 & & \\ & s_2 & \\ & & 1 \end{bmatrix} \begin{bmatrix} \mathsf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \mathbf{x} \qquad (1)$$

where R is a rotation matrix such that $\mathsf{R}^\top\mathsf{R} = \mathsf{R}\mathsf{R}^\top = \mathsf{I}$ and $\mathbf{t}$ is a translation vector. s_1 and s_2 are scale factors which are independent along to the orthogonal axis.

Note that eq.(1) is a kind of affine transformations and looks similar to the metric transformation except the orthogonal scale matrix. In the semi-metric space, the metric properties along the parallel lines which are aligned to the X,Y axis in the warped plane are all preserved, but ones not aligned to the axis are not preserved. Of course, affine properties are all preserved in the warped plane, because it is one of affine transformations.

Strictly speaking, the semi-metric transformation cannot be a general stratification of projective transformation because of these properties. However, the semi-metric space is one of good tools to analyze scenes that have only information about parallelism and orthogonality of some planes.

2.1 Warping to the semi-metric space

For metric rectification of a projective distorted plane, there are some ways to find the rectifying homographys [2, 3]. Generally it is possible with five independent orthogonal line sets, or with a rectangle whose aspect ratio is known, or with a line at infinity and an orthogonal line set. Essentially, these three conditions are all equivalent [2] to the case with a rectangle whose aspect ratio is known. If we do not have the sufficient condition, how can we warp the plane into the semi-metric space?

There are two ways to warp an image to a semi-metric space. First one is using orthogonal vanishing points, and the other is using a standard rectangle.

2.1.1 Using orthogonal vanishing points

As derived in [4], a projected dual circle is given as

$$\mathsf{A}^{-1} = \mathbf{\Delta}_2 + s_3\mathbf{x}_c\mathbf{x}_c^\top$$

and it is expressed as

$$\begin{aligned}\mathsf{A}^{-1} &= s_1\mathbf{v}_1\mathbf{v}_1^\top + s_2\mathbf{v}_2\mathbf{v}_2^\top + s_3\mathbf{x}_c\mathbf{x}_c^\top \\ &= \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{x}_c\end{bmatrix} \operatorname{diag}(s_1, s_2, s_3) \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{x}_c\end{bmatrix}^\top \\ &\triangleq \mathsf{V}\mathsf{D}\mathsf{V}^\top\end{aligned}$$

where D is a diagonal scale matrix and V is a matrix which contains orthogonal vanishing points $\mathbf{v}_1$, $\mathbf{v}_2$ and an origin of the target plane $\mathbf{x}_c$. Assume that the plane homography is P. Without loss of generality, the matrix V is expressed as

$$\mathsf{V} = \mathsf{P}\operatorname{diag}(a, b, c) \qquad (2)$$

where a, b and c are proper scale factors that are needed to correct the scales. This means that warping with matrix V^{-1} makes planes independently scaled along to the orthogonal axis, and this is a *semi-metric image*. The resulting warping matrix is V^{-1}.

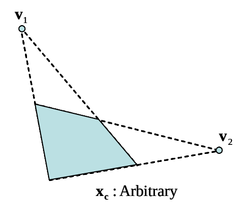

Fig. 1: Elements of semi-metric transformation matrix from vanishing points

This formulation is similar to other vanishing point based algorithms [5, 6]. The key difference is the third column of the matrix V. Previously, they set it with the third orthogonal vanishing points, but here, it is set with an arbitrary point on the plane.

2.1.2 Using a standard rectangle

Finding vanishing points from a projected rectangle is not easy. To warp an image to semi-metric space, there is the other option to use the projected rectangle itself. A warping from the projected rectangle to a standard rectangle is sufficient to build a semi-metric space. A standard rectangle is a predefined rectangle whose aspect ratio is known. Fig.2 shows the concept of the warping method using a standard rectangle.

The warping matrix H_{sm} is computed by a conventional plane homography estimation algorithm using four points [3]. Due to some numerical issues, a normalized algorithm with a standard rectangle which is similar to the algorithm proposed by Hartley [7] is preferred.

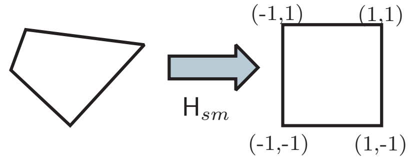

Fig. 2: Semi-metric warping using a standard rectangle

The matrix H_{sm} is equivalent to a warping matrix V using vanishing points, because the resulting warped images from H_{sm} and one from V can be transferred to each other by translating and orthogonal scaling.

3 About off-the-plane features

In this section, we study about features which are not on the model plane. If we warp a projectively distorted plane to a semi-metric space, there are also some interesting properties regarding the off-the-plane features. The properties are generally known as *plane+parallax properties* [3, 8–11]. However, by using semi-metric warping, a simpler derivation is possible.

Also we can make a *semi-metric projection matrix* from the semi-metric warped image. The semi-metric 3D structure is obtained from a single image of a projectively distorted scene with some additional knowledge of the scene, which is similar to the method proposed in [12,13] without three strong vanishing points.

3.1 Points off the reference plane in semi-metric space

Because the original image is fully transformed projectively, there are some feature points that are not on the reference plane. Although the semi-metric warping is achieved, a projective distortion along to the third orthogonal direction is remaining. In this section, we investigate the position of the off-the-plane points after the semi-metric warping.

Without loss of generality, we use a matrix V^{-1} as a semi-metric warping transformation. By the general pin-hole projection model, a projected point of an obtained feature point X whose coordinates are (X, Y, Z) is given as

$$\mathbf{x} \sim \mathsf{K}\begin{bmatrix}\mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t}\end{bmatrix}\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix} \tag{3}$$

where K denotes a camera matrix describing internal parameters of the camera.

By warping $\mathbf{x}$ in eq.(3) and V^{-1} in eq.(2) to the semi-metric space, the warped point $\mathbf{x}'$ is

$$\mathbf{x}' \sim \mathsf{V}^{-1}\mathbf{x}. \tag{4}$$

Eq.(4) can be rewritten as

$$\mathbf{x}' = \begin{bmatrix}c/a & & \\ & c/b & \\ & & 1\end{bmatrix}\begin{bmatrix}X + Zm_1\\Y + Zm_2\\1 + Zm_3\end{bmatrix}, \tag{5}$$

where $\mathbf{m}$ is defined as $\mathbf{m} \triangleq \begin{bmatrix}\mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t}\end{bmatrix}^{-1}\mathbf{r}_3 = \begin{bmatrix}m_1 & m_2 & m_3\end{bmatrix}^\top$. Using inhomogeneous coordinate, the point in the semi-metric spcae is expressed as

$$\left(\frac{c}{a}\frac{X + Zm_1}{1 + Zm_3}, \frac{c}{b}\frac{Y + Zm_2}{1 + Zm_3}\right) \tag{6}$$

Note that the scale factors c/a and c/b can be canceled out because the point is in the *semi-metric space*.

3.2 Relative-Z estimation

Based on the eq. (6) , we can extract some useful information of the scene.

Assume that there are two points $(X_0, Y_0, 0)$ and $(-X_0, -Y_0, 0)$ on the reference plane. A length between the two points on the semi-metric space is calculated from eq.(6) as

$$l_0 = \alpha_1 \frac{c}{a} 2X_0 \tag{7}$$

where α_1 is a scale factor to describe an arbitrary semi-metric warping. In the same semi-metric space, a difference of X direction in the semi-metric space between two equal-Z points (X_1, Y_1, Z) and (X_2, Y_2, Z) is given by

$$l = \alpha_1 \frac{c}{a}\frac{X_1 - X_2}{1 + Zm_3}. \tag{8}$$

From eqs.(7) and (8), the length ratio is given as

$$\frac{l}{l_0} = \frac{1}{1 + Zm_3}\frac{L}{L_0}$$

where L is a difference of X coordinate between two points which have common Z coordinate, and L_0 is a difference of X coordinate between two points on the reference plane. Some tedious manipulations of the equation make

$$\frac{1}{1 + Zm_3} = \left(\frac{l}{l_0}\Big/\frac{L}{L_0}\right) \triangleq L'.$$

So the relative Z, that is Zm_3, is calculated easily as

$$Zm_3 = \frac{1 - L'}{L'}.$$

It is not necessary to know m_3, because this equation gives us only *relative-Z* coordinate, if we know the ratio of differences of X or Y coordinates of the two line segments in metric or Euclidean space. Generally these values are easily obtained through a semi-metric warping of a target plane using the property of a semi-metric space that conserves the length ratio along to the orthogonal axis.

3.3 Semi-metric projection matrix

Once we find the information about relative-Z's of scene features, we can define a *semi-metric 3D space* of the scene. We can derive a formulation about the 3D-2D relationship of semi-metric space. We call it a *semi-metric projection matrix*.

Eq.(6) can be rewritten as

$$\begin{bmatrix} x_{sm} \\ y_{sm} \\ 1 \end{bmatrix} \sim \begin{bmatrix} \frac{1}{a} & & \\ & \frac{1}{b} & \\ & & \frac{1}{c} \end{bmatrix} \begin{bmatrix} 1 & & m_1 & \\ & 1 & m_2 & \\ & & m_3 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix}. \tag{9}$$

Here, the leftmost matrix $\mathrm{diag}(1/a, 1/b, 1/c)$ is one of semi-metric transformation matrices. Therefore we define a pure semi-metric projection matrix as

$$\begin{bmatrix} 1 & & m_1 & \\ & 1 & m_2 & \\ & & m_3 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{m} & \mathbf{e}_3 \end{bmatrix}$$

where $\mathbf{e}_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^\top$, $\mathbf{e}_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^\top$ and $\mathbf{e}_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^\top$.

To complete the semi-metric projection matrix, the only remaining problem is to compute a vector $\mathbf{m}$.

3.3.1 Physical meaning of m

From eq.(5), $\mathbf{m}$ is defined as

$$\mathbf{m} = \begin{bmatrix} m_1 & m_2 & m_3 \end{bmatrix}^\top = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}^{-1} \mathbf{r}_3.$$

Assume that $\mathbf{m}$ is a point in a semi-metric space. If we try to warp the point into a projective space with semi-metric warping matrix V^{-1}, the point $\mathbf{m}'$ is given as

$$\mathbf{m}' = \mathsf{V}\mathbf{m} = \mathsf{K}\mathbf{r}_3 = s\mathbf{v}_{sm},$$

where $\mathbf{v}_{sm}$ is a vanishing point whose direction is orthogonal to the reference plane in the projective space, that is *the third vanishing point*.

We can conclude that the vector $\mathbf{m}$ is the scaled third vanishing point in a semi-metric space. If we know the position of the vanishing point in a semi-metric image, we have to find a proper scale factor to complete the semi-metric projection matrix.

3.3.2 Computing m

If we know the relative-Z's with respect to the reference plane, we can calculate the X and Y coordinates of the point in the semi-metric space from eq.(9). Some basic manipulations with eq.(9) give us

$$\begin{aligned} X &= x_{sm}(1 + m_3) - m_1 Z \\ Y &= y_{sm}(1 + m_3) - m_2 Z. \end{aligned} \tag{10}$$

We can make two algorithms to compute $\mathbf{m}$ from eq.(10). The first one is to obtain the scale s if we know the orthogonal vanishing point $\mathbf{v}_{sm}$ in the semi-metric space. The second one is to find the vector $\mathbf{m}$ from some scene constraints directly.

1) Algorithm with known orthogonal vanishing point $\mathbf{v}_{sm}$

Assume that there are two points which have the same X coordinate in Euclidean world. Because $\mathbf{m} = s\mathbf{v}_{sm}$ and $X = x_{sm}(1 + m_3) - m_1 Z$, for the two points with the same X coordinate, we obtain

$$X = x_{sm1}(1+m_3) - m_1 Z_1 = x_{sm2}(1+m_3) - m_1 Z_2.$$

Therefore, we can derive the scale s from the equation as

$$s = \frac{x_{sm2} - x_{sm1}}{v_{sm_x}(Z_2 - Z_1) + (x_{sm1} Z_1 - x_{sm2} Z_2)}$$

where the orthogonal vanishing point $\mathbf{v}_{sm} = \begin{bmatrix} v_{sm_x} & v_{sm_y} & 1 \end{bmatrix}^\top$.

If it is the case that two Y coordinates of the points are equal, the same derivation is possible. The required constraint is that there are two points which have the same X or Y coordinate in Euclidean (or metric) world. This condition is occurred often in treating an image of real worlds.

2) Algorithm to compute m directly from some scene constraints

If we have more pairs of points that have common X and Y coordinates, it is possible to estimate $\mathbf{m}$ with a proper scale directly. We can make basic equations as

$$X = x_{sm1}(1+m_3) - m_1 Z_1 = x_{sm2}(1+m_3) - m_1 Z_2$$

and

$$Y = y_{sm1}(1+m_3) - m_2 Z_1 = y_{sm2}(1+m_3) - m_2 Z_2.$$

These equations are rewritten as

$$(-Z_1+Z_2)m_1 + (x_{sm1}Z_1 - x_{sm2}Z_2)m_3 = x_{sm2} - x_{sm1}$$

$$(-Z_1+Z_2)m_2 + (y_{sm1}Z_1 - y_{sm2}Z_2)m_3 = y_{sm2} - y_{sm1}.$$

You can notice that it needs one more equation to make it over-constrained. Therefore the minimal condition is two pairs of points which have the same X and Y coordinates. Of course, pairs of points which have only one common X or Y coordinates, not both, are helpful to make the equation over-constrained.

3.4 Semi-metric 3D reconstruction with a single view

Using the methods described in the previous sections, we can build a semi-metric 3D model with a single view. An algorithm is summarized as follow.

1. Warping to a semi-metric space

   Using methods described in section 2.1, warping to a semi-metric space is possible. Tracking of the orthogonal vanishing points or tracking of rectangular features is sufficient to warp an image to a semi-metric space.

2. Computing relative-Z's

   Relative Z, that is a ratio of the orthogonal distances from the reference plane, can be extracted by a method represented in section 3.2. To compute relative Z of the features, ratios of physical length of line segments which are parallel to the reference plane are needed, and it can be obtained from setting by humans or from another semi-metric warping with respect to a plane which contains the line segments.

3. Estimating semi-metric projection matrix

   For this, we can utilize the methods in section 3.3. Essentially, the objective of this part is to obtain a third vanishing point whose direction is orthogonal to the reference plane and to find a proper scale factor. We only need some pairs of points that have common X or Y coordinates and their relative-Z's.

4. Computing X and Y coordinates

   From eq.(10), X and Y coordinates of feature points are obtained directly. Reversely, if we know X or Y coordinates of feature points, a proper relative-Z can be estimated from eq.(10), which is similar with the method proposed in [13].

5. Constructing a semi-metric 3D structure

   Construction of a semi-metric 3D structure is possible by linking features topologically. The topology, which means the connection between features, is determined manually using prior knowledge about the scene.

3.4.1 Experiment with a real image

In this section, we show an example of reconstruction of semi-metric 3D with a real image. Fig. 3 is an input image of a building scene. The image is captured using a camera module attached in SAMSUNG SPH-2500 mobile phone, whose intrinsic parameters cannot be adjusted. There are several artificial planes, and we can easily detect parallel line segments in the image. We selected three planes in the image to be reconstructed.

Fig. 3: Real input image for verification of proposed algorithm

Fig. 4 shows a reconstructed 3D structure in semi-metric 3D space. Note that the lines along to the orthogonal axis are all orthogonal to each other, although we didn't apply any kind of robust methods for estimating seme-metric 3D reconstruction. We used only easily-obtainable information of the scene, for example, a rectangle, parallel lines to find the orthogonal vanishing points, and constraints for line segments whose length are all equal. Note that there were no extrinsic measurements in the process.

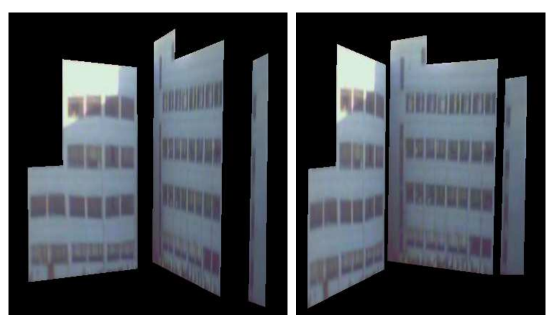

Fig. 4: Reconstructed semi-metric 3D structure (VRML) model

4 Conclusion

In this paper, we propose a new method to reconstruct 3D structures of captured scenes using parallelism and orthogonality. For this, we propose a new transformation group named as *semi-metric space* and we reveal the properties of the space and its benefits. In semi-metric space, the metric properties aligned to each predefined orthogonal axis are preserved, but ones not aligned to predefined axis are not metric invariants. If we have some information of parallelism and orthogonality, upgrade to the semi-metric space can be simply achieved. From the semi-metric space, the partial structure of the scene can be retrieved from a single image and its easily-obtainable scene constraints. The resulting 3D is called semi-metric 3D, and we can find the scene structure up to semi-metric transformation even if we have only one image and no external measurements.

Acknowledgment This research has been supported by the Korean Ministry of Science and Technology for National Research Laboratory Program(Grant number M1-0302-00-0064).

References

[1] S. Palmer, *Vision Science: Photons to Phenomenology.* MIT Press, 1999.

[2] D. Liebowitz, *Camera Calibration and Reconstruction of Geometry from Images.* PhD thesis, University of Oxford, 2001.

[3] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision.* Cambridge, 2000.

[4] J.-S. Kim, P. Gurdjos, and I. S. Kweon, "Geometric and algebratic constraints of projective concentric circles and their applications to camera calibration," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2004. Accepted.

[5] B. Caprile and V. Torre, "Using vanishing points for camera calibration," *International Journal of Computer Vision*, vol. 4, pp. 127–140, March 1990.

[6] R. Cipolla, T. Drummond, and D. Robertson, "Camera calibration from vanishing points in images of architectural scenes," in *British Machine Vision Conference*, 1999.

[7] R. Hartley, "In defense of the eight-point algorithm," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, pp. 580–593, June 1997.

[8] A. Criminisi, I. Reid, and A. Zisserman, "Duality, rigidity and planar parallax," in *European Conference on Computer Vision*, p. II: 846, 1998.

[9] C. Rother and S. Carlsson, "Linear multi view reconstruction and camera recovery using a reference plane," *International Journal of Computer Vision*, vol. 49, pp. 117–141, September 2002.

[10] M. Irani, P. Anandan, and D. Weinshall, "From reference frames to reference planes: Multi-view parallax geometry and applications," in *European Conference on Computer Vision*, p. II: 829, 1998.

[11] M. Irani, P. Anandan, and M. Cohen, "Direct recovery of planar-parallax from multiple frames," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1528–1534, November 2002.

[12] A. Criminisi, I. Reid, and A. Zisserman, "A plane measuring device," *Image and Vision Computing*, vol. 17, pp. 625–634, June 1999.

[13] A. Criminisi, I. Reid, and A. Zisserman, "Single view metrology," *International Journal of Computer Vision*, vol. 40, pp. 123–148, November 2000.

Kim, Jun-Sik: received the B.S. degree in Electrical Engineering from Yonsei University in 1999 and the M.S. degree in Electrical Engineering and Computer Science from KAIST, Korea in 2001. He is now pursuing Ph.D. degree in Electrical Engineering and Computer Science at KAIST. His research interest includes computer vision, camera calibration, robotics, photogrammetry and 3D reconstruction.

Kweon, In So: received the B.S. and the M.S. degrees in Mechanical Design and Production Engineering from Seoul National University, Seoul, Korea, in 1981 and 1983, respectively, and the M.S. and Ph.D. degrees in Robotics from the Robotics Institute at Carnegie Mellon University, Pittsburgh, U.S.A, in 1986 and 1990,respectively. He worked for Toshiba R&D Center, Japan, and joined the Department of Automation and Design Engineering at KAIST in 1992. He is now a Professor in the Department of Electrical Engineering and Computer Science at KAIST. His research interests are in computer vision, robotics, pattern recognition, and automation. Specific research topics include invariant based visions for recognition and assembly, 3D sensors and range data analysis, color modeling and analysis, robust edge detection, and moving object segmentation and tracking.