METRIC PLANE RECTIFICATION USING SYMMETRIC VANISHING POINTS

M. Lefler[*], *H. Hel-Or*

Dept. of CS, University of Haifa, Israel

Y. Hel-Or

School of CS, IDC, Herzliya, Israel

ABSTRACT

Video analysis often requires mapping of activity or object locations from image coordinates to ground plane coordinates. This process is termed *Plane Rectification*. In this paper we propose an geometric method to find plane rectification using the plane's vanishing line and the vertical vanishing point. Unlike common methods that provide sophisticated algebraic solutions and non-linear optimizations, the proposed approach is intuitive and simple to implement while providing a geometric explanation and interpretation of the plane rectification. We show that the proposed approach provides stable and accurate solutions also in the presence of noise.

***Index Terms*—** Plane Rectification, symmetric vanishing points, people tracking, video sequences.

1. INTRODUCTION

Video surveillance cameras are often configured such that tracking, trajectory analysis and multi view merging require mapping multiple images into a unified frame of reference. Having a ground plane in the scene suggests the use of *Plane Rectification* which transforms the images to a view as if they were acquired by an orthographic projection where the projection plane is parallel to the ground plane. In fact, this can be seen as a transformation from image coordinates to the 2D ground plane coordinates. We distinguish between *affine rectification* where parallelism is retained, but other properties such as angle size and ratio of lengths are not, and *metric rectification* where distances are preserved up to similarity (translation, rotation and isotropic scale).

In contrast to existing methods for metric plane rectification which use algebraic constraints and complex entities, we present here a rectification method which is purely geometric. The geometry aspect of the approach, provides a novel point of view for the rectification process, making it easier to visualize and understand.

In this paper, we define a a set of three special vanishing points which we term *symmetric vanishing points*. These points are extracted from the the ground plane vanishing line and the vertical vanishing point. We show that a metric rectification matrix can be composed directly from these symmetric points with no additional computations.

[*]supported by the Chief Scientist Office at the Israeli Ministry of Industry, Trade & Labor, under the Magnet Program (VULCAN).

2. DIRECT METRIC RECTIFICATION

Consider a world scene with a ground plane acquired by a (pinhole) camera (Figure 1). We define a world coordinate system XYZ, such that the origin of the coordinate system lies on the ground plane, the Z axis is perpendicular to the plane, and the X and Y axes span the ground plane. The transformation from world coordinates to image coordinates can be simply defined as

$$\mathbf{x} = P\mathbf{X} \tag{1}$$

where $\mathbf{X} = (X, Y, Z, 1)^T$ is a homogeneous 4-vector representing a point in the world, P is the 3×4 homogeneous camera projection matrix, and $\mathbf{x} = (x, y, w)^T$ denotes a homogeneous 3-vector representing the projected image point. Restricting the projection to the ground plane alone, $(Z = 0)$, we consider only world points of the form $\mathbf{X} = (X, Y, 0, 1)$. The projection can then be reduced to a 3×3 homography H that projects points on the world's ground plane to the image plane. Inverting this homography gives the rectification matrix that transforms the image back to the world plane:

$$M = H^{-1}$$

The homography H can be decomposed into a similarity transformation H_S, an affine transformation H_A and a projective transformation H_P [1]:

$$H = H_P H_A H_S$$

H_P can be constructed from the image coordinates of two vanishing points $\mathbf{v_1}$, $\mathbf{v_2}$ on the vanishing line of the ground plane, and the image coordinates $\mathbf{v_3}$ of the world origin [2]:

$$H_P = \begin{pmatrix} v_{1_x} & v_{2_x} & v_{3_x} \\ v_{1_y} & v_{2_y} & v_{3_y} \\ v_{1_w} & v_{2_w} & v_{3_w} \end{pmatrix} \tag{2}$$

Image rectification using $M = H_P^{-1}$, is an affine rectification. In order to obtain a metric rectification additional computations must be performed, e.g. evaluation of the calibration parameters (external camera parameters) as in [2],[3].

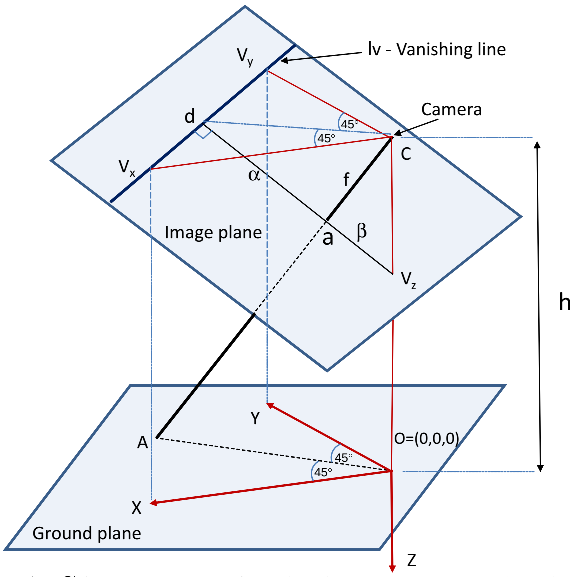

Fig. 1: C is the center of projection (camera), a the piercing point and A the intersection of the optical axis with the ground plane. The two symmetric vanishing points $\mathbf{v_x}$ and $\mathbf{v_y}$ are symmetric about d on the ground plane's vanishing line.

In this paper ,we show that by carefully selecting three orthogonal vanishing points - the *symmetric vanishing points* - and collecting them in a matrix form as in Equation 2, a **metric** rectification can be obtained.

3. RECTIFICATION USING SYMMETRIC VANISHING POINTS

Assume a vanishing line $\mathbf{l_v}$ associated with the world plane, and a vertical vanishing point $\mathbf{v_z}$ associated with the direction perpendicular to the world plane are given. Any method may be used to compute these. For example, in [4, 5] trajectories of people moving on a planar ground induce head and feet parallel lines from which the vanishing line is computed. The vertical vanishing point $\mathbf{v_z}$ is computed from the intersection of head-feet lines. In [3], parallel lines in an indoor scene are found from which 3 orthogonal vanishing points are obtained. Two of these are on the vanishing line and the third is $\mathbf{v_z}$. In [2], known lines associated with house walls are used to find vanishing points from which $\mathbf{v_z}$ and $\mathbf{l_v}$ can be extracted.

We define the *symmetric vanishing points* as $\mathbf{v_x}$ and $\mathbf{v_y}$, two vanishing points on $\mathbf{l}_v$, such that, together with $\mathbf{v_z}$, are associated with an orthogonal triplet of directions in 3D. The uniqueness of the points $\mathbf{v}_x$ and $\mathbf{v}_y$ is that they are symmetric with respect to the image line $(\mathbf{v}_z, \mathbf{d})$ passing through $\mathbf{v}_z$ and perpendicular to $\mathbf{l}_v$ (Figure 1).

Collecting the symmetric vanishing points (in homogeneous coordinates) in columns, we show that:

$$M = H_m^{-1} = \begin{pmatrix} v_{x_x} & v_{y_x} & v_{z_x} \\ v_{x_y} & v_{y_y} & v_{z_y} \\ 1 & 1 & 1 \end{pmatrix}^{-1} \tag{3}$$

is a metric rectification matrix.

Proof. Consider the scene and camera setup shown in Figure 1. C is the center of projection and a the image piercing point. We define the origin O of the world coordinate system to be at the perpendicular projection of C on the world plane. The world coordinate axes X,Y and Z are the three mutually orthogonal directions induced by the three symmetric vanishing points: $\mathbf{v_x}$, $\mathbf{v_y}$, and $\mathbf{v_z}$. Given a 3D point $\mathbf{X} = (X, Y, Z, 1)$, its image $\mathbf{x} = (x, y, w)$ in the image plane is given by Eq. 1. The projection matrix P is decomposed as [1]:

$$P = KR\,[I;\ -C] = [KR;\ -KRC] \tag{4}$$

where C is the camera's center in world coordinates and R is a rotation matrix of the world coordinate system to camera coordinate system. Assuming a non-skew camera and piercing point a at the image's center we have:

$$K = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5}$$

is the camera's internal parameter matrix with focal length f. Assuming the camera is at height h above the world ground plane, we have $C{=}(0, 0, -h)$ and the camera z-axis $A - C = (A_x, A_y, 0 + h)$ where A is the intersection of the z-axis and the ground plane. This selection of the world axes yields that the ray (O, A) on the ground plane is projected on the image plane to the line $(\mathbf{v_z}, a)$, which intersects the vanishing line $\mathbf{l_v}$ at point d. It can easily be seen that the symmetric selection of vanishing points about d yields that the world axes are also symmetric about (O, A) and that the world X and Y axes both form a 45 degree angle with respect to the direction (O, A). Thus, the representation of point A in world coordinates is $A_x = A_y$ and $A = (A_x, A_x, 0)$. The camera's z-axis is then:

$$\mathbf{R}_z = A - C = (A_x, A_x, h) \tag{6}$$

Now consider again, the camera projection matrix P (Eq. 4). The rotation matrix R is given by composing the camera's axes vectors $\mathbf{R}_x, \mathbf{R}_y, \mathbf{R}_z$ in matrix form. With Eq. 6 and 5 and given that $C = (0, 0, -h)$, we obtain:

$$P = [KR;\ -KRC] = \begin{pmatrix} R_{x_1} f & R_{x_2} f & R_{x_3} f & R_{x_3} fh \\ R_{y_1} f & R_{y_2} f & R_{y_3} f & R_{y_3} fh \\ A_x & A_x & h & h^2 \end{pmatrix} \tag{7}$$

The world coordinates $(1, 0, 0, 0)$,$(0, 1, 0, 0)$,$(0, 0, 1, 0)$ and $(0, 0, 0, 1)$, project onto the image points $\mathbf{v_x}$,$\mathbf{v_y}$,$\mathbf{v_z}$,$\mathbf{v_z}$ respectively. However due to homogeneity of the coordinates, these image points are unique up to a scale factor λ. Thus:

$$P = \begin{pmatrix} v_{x_x} & v_{y_x} & v_{z_x} & v_{z_x} \\ v_{x_y} & v_{y_y} & v_{z_y} & v_{z_y} \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix} \tag{8}$$

To obtain the homography, H that maps the world plane to the image plane, we omit the third column of P (since ground plane points have z-coordinate equal 0):

$$H = \begin{pmatrix} v_{x_x} & v_{y_x} & v_{z_x} \\ v_{x_y} & v_{y_y} & v_{z_y} \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_4 \end{pmatrix} \tag{9}$$

From 7 and 8 we have:

$$\lambda_1 = A_x \quad \lambda_2 = A_x \quad \lambda_4 = h^2$$

Thus, $\lambda_1 = \lambda_2$ and defining λ_4 as the homogeneous scale factor and setting $\lambda = \frac{\lambda_1}{\lambda_4}$, we obtain:

$$H = \begin{pmatrix} v_{x_x} & v_{y_x} & v_{z_x} \\ v_{x_y} & v_{y_y} & v_{z_y} \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix} = H_m H_S \tag{10}$$

with H_S being a similarity transformation (uniform scale). Thus it can be deduced that

$$H_m = \begin{pmatrix} v_{x_x} & v_{y_x} & v_{z_x} \\ v_{x_y} & v_{y_y} & v_{z_y} \\ 1 & 1 & 1 \end{pmatrix} \tag{11}$$

and

$$M = H_m^{-1} \tag{12}$$

is a metric rectification matrix. Note that any other world coordinate system (with Z axis perpendicular to the world plane) is a similarity transformation away, thus maintaining that M is a metric rectification (i.e. up to similarity) of any other world coordinate system. Also note that if any other triplet of orthogonal vanishing points is selected, the constraint $\lambda_1 = \lambda_2$ does not hold in Eq. 9 which results in M being an affine rather than metric rectification matrix. $\square$

3.1. Rectification in Practice

In order to compute the symmetric vanishing points, the camera's focal length f must be computed. Since (C, d) is perpendicular to $(C, \mathbf{v_z})$ [1], simple geometry shows that the triangles $(C, a, \mathbf{v_z})$ and (d, a, C) are similar. Thus, defining $\alpha = \|(d, a)\|$ and $\beta = \|(a, \mathbf{v_z})\|$, it is easy to show that:

$$f = \sqrt{\alpha\beta} \tag{13}$$

Now, since (C, d) is mid-perpendicular to $(\mathbf{v_x}, \mathbf{v_y})$ (it is on the plane spanned by $(C, \mathbf{v_z})$ and (C, a), which is mid-perpendicular to the symmetric vanishing points by definition) and its length is $\sqrt{\alpha^2 + f^2}$, we have

$$\|(\mathbf{v_x}, d)\| = \|(\mathbf{v_y}, d)\| = \sqrt{\alpha^2 + f^2} \tag{14}$$

The following algorithm summarizes the process of computing the metric rectification matrix (refer to Figure 1).
Given the vanishing line $\mathbf{l_v}$ and vanishing point $\mathbf{v_z}$:

1. Determine the point d on the vanishing line which is the perpendicular projection of the vanishing point $\mathbf{v_z}$ onto the vanishing line.
2. Compute the camera focal length: Define $\alpha = \|(d, a)\|$ and $\beta = \|(a, \mathbf{v_z})\|$, and compute $f = \sqrt{\alpha\beta}$. Note, that in real images, the piercing point often does not necessarily align with image center. We follow [4] to correct for this.
3. Determine the symmetric vanishing points $\mathbf{v_x}$ and $\mathbf{v_y}$ on the vanishing line that are on either side of d at the distance: $\|(\mathbf{v_x}, d)\| = \|(\mathbf{v_y}, d)\| = \sqrt{\alpha^2 + f^2}$
4. Collect the three vanishing point coordinates (in normalized homogeneous coordinates) in the matrix H:

$$H = \begin{pmatrix} v_{x_x} & v_{y_x} & v_{z_x} \\ v_{x_y} & v_{y_y} & v_{z_y} \\ 1 & 1 & 1 \end{pmatrix} \tag{15}$$

5. Define the Metric rectification M:

$$M = H^{-1}$$

3.2. Why it works - Insights

In principle, given the vertical vanishing point $\mathbf{v_z}$, any 2 orthogonal points on the vanishing line $\mathbf{l_v}$ can be used to form a non-metric rectification matrix (Eq. 2). However to create a metric rectification, the skew factors λ_1 and λ_2 (Eq. 9) must be determined whether explicitly [3] or implicitly [2, 5]. By choosing the unique configuration, using the symmetric vanishing points, there is no skew as $\lambda_1 = \lambda_2$ and there is no need to correct the transformation for these values.

A geometric interpretation of this key point is as follows (see Figure 1). Image rectification can be viewed as involving a 3D transformation that maps the image plane (in 3D) to the world plane (in 3D). Without loss of generality assume the transformation involves a rotation about the 3D camera position C so that the image plane is parallel with the world plane, followed by a 3D translation that aligns the 2 planes. Further assume the rotation is a rotation about the camera's pitch axis which passes through point C and parallel to the vanishing line (degree of rotation is given by the angle $\measuredangle(\mathbf{v_z}, C, a)$)[1]. Now, given any 2 orthogonal vanishing points $\mathbf{v_1}, \mathbf{v_2}$, this pitch will map the vanishing points to the world plane such their distance from the origin O is unequal and proportional to $\|(\mathbf{v_z}, \mathbf{v_1})\| / \|(\mathbf{v_z}, \mathbf{v_2})\|$. This ratio is the source of skew in the final rectification and can be shown to equal λ_1/λ_2! Thus to eliminate the skew, one must choose 2 vanishing points on $\mathbf{l_v}$ that are symmetric about the plane perpendicular to the pitch angle and passing through C - namely 2 vanishing points that are symmetric about the point d in the image.

[1] Note that any other transformation that performs the plane alignment can be shown to be an in-plane similarity transformation of the one assumed here.

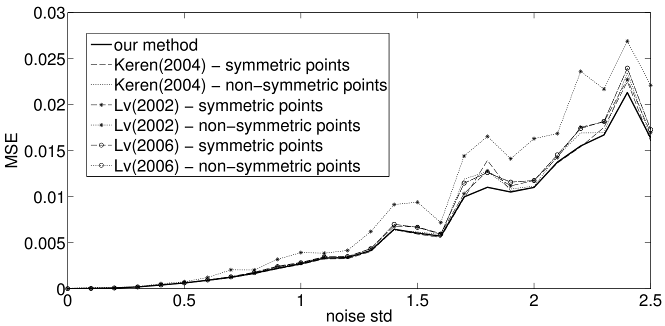

Fig. 2: Comparison of results run on synthetic data.

4. RESULTS

We show that the suggested approach using symmetric vanishing points improves on other approaches in terms of error rate. Additionally, we show the advantage of replacing the original vanishing points with the symmetric vanishing points in other methods. We note, however, that other approaches require calculating the external camera parameters regardless of the vanishing points used.

Testing was performed on image sequences of people walking on a ground plane. The method described in [4] was used to compute the vertical vanishing point and to compute numerous vanishing points on the vanishing line using tracked heads and feet of the people. The calculated vanishing line and vertical vanishing point were used as input to the algorithms tested in this section.

Results from synthetic data - A synthetic world was created and this configuration was later used as ground truth. Targets of different heights (simulating people) were placed randomly on the world plane. These targets then "moved" in random directions and distances creating a second "time frame" of the world. The two world time frames were projected onto an image plane using a virtual camera with preset parameters ($h = 80$, $f = 10$ and $A = (50, 100, 0)$ - see Figure 1). Head and feet coordinates were contaminated with Gaussian noise resulting in noisy vertical vanishing point and noisy vanishing line. Tests were run at various noise levels (std=0 .. 2.5 measured in image pixel units). 1000 trials of different initial data point configurations were tested per each noise level. For each trial, image data points were rectified back to the world plane using four different methods: The method presented in this paper, and the methods presented in[4], [5] and [3]. For fair comparison, we did not implement the post-processing optimizations suggested in [5]. Note, that in Euclidean rectification algorithms a height parameter is required. Since we are interested only in metric rectification, an arbitrary height was provided (which, for these algorithms, affects unit scale alone). Each method was tested using symmetric vanishing points, which were computed as described in this paper, and randomly generated non-symmetric orthogonal vanishing points, computed following [4]. We compared the results for each rectification method by finding the best similarity transformation [6] that maps the rectified points to the ground truth, and computing the mean square error of the mapped points against the ground truth. Results are shown in Figure 2. Results show that the proposed method performs better than the other approaches. Furthermore, using symmetric points in other methods (vs. using vanishing points as originally suggested in these methods) improves performance.

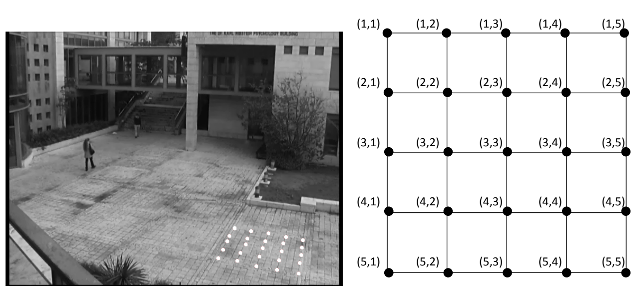

Fig. 3: (a) A frame of the video sequence with tracked person and selected data points (in white). (b) Ground truth pattern associated with the selected data points in a.

	Proposed method	Keren *et al.* (2004)	Lv *et al.* (2002)	Lv *et al.* (2006)
S	2.015	2.021	2.017	2.07
NS		m = 2.194 std = 0.501	m = 2.018 std = 0.088	m = 2.022 std = 0.128

Table 1: Results on real data (S=symmetric NS=non-symmetric vanishing points).

Results from real data - Testing was performed on a real scenario using a video sequence of walking people (Figure 3a). People were tracked using the tracker from [7] and the method described in [4] was used to compute the vanishing line and vertical vanishing point. A grid of 25 points marked in the scene served as the ground truth (see Figure 3b). The image points of the ground truth grid were rectified using the same set of methods as above. The rectified points were compared to the ground truth up to similarity as above. Results are shown in Table 1. The results once again confirm that the proposed method outperforms other tested methods and that using symmetric vanishing points rather than any other vanishing point improves the results in the other methods as well.

5. CONCLUSION

An intuitive and simple method for metric plane rectification is suggested. We obtain metric rectification without the need to calculate camera parameters or extract any other features or properties from the scene. The method computes the *symmetric vanishing points* which are collected in a matrix form similar to the approach used for affine rectification. However, due to the characteristics of the symmetric points the resulting rectification is metric. Experiments on simulated and real data validated the advantage of the proposed method.

6. REFERENCES

[1] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge Univ Pr, 2003.

[2] D. Liebowitz, A. Criminisi, and A. Zisserman, "Creating architectural models from images," in *Computer Graphics Forum*, 1999, vol. 18, pp. 39–50.

[3] S. Keren, I. Shimshoni, and A. Tal, "Placing three-dimensional models in an uncalibrated single image of an architectural scene," *Presence: Teleoperators & Virtual Environments*, vol. 13, no. 6, pp. 692–707, 2004.

[4] F. Lv, T. Zhao, and R. Nevatia, "Self-calibration of a camera from video of a walking human," in *International Conference on Pattern Recognition*, 2002, vol. 16, pp. 562–567.

[5] F. Lv, T. Zhao, and R. Nevatia, "Camera calibration from video of a walking human," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 9, pp. 1513, 2006.

[6] S. Umeyama, "Least-squares estimation of transformation parameters between twopoint patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 376–380, 1991.

[7] D. Ross, J. Lim, R.S. Lin, and M.H. Yang, "Incremental Learning for Robust Visual Tracking," *International Journal of Computer Vision, Special Issue: Learning for Vision*, vol. 77, no. 1-3, pp. 125–141, 2007.