Ground Plane Rectification by Tracking Moving Objects

Biswajit Bose and Eric Grimson

Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139, USA

Abstract

Most outdoor visual surveillance scenes involve objects moving on a ground plane. We present a new, fully automated technique for both affine and metric rectification of this ground plane (up to a scale factor) by simply tracking moving objects. In particular, we derive the necessary constraints on the image plane to ground plane projective transformation by observing objects which move at constant (world) velocity for some part of their trajectory. No knowledge of camera parameters is assumed. We describe a hierarchy of possible solutions, depending on the nature of the motion trajectories observed. We also show how to automatically detect degenerate cases where 2D rectification is not possible. Useful applications of the various types of rectification are presented. Our experiments demonstrate all the possible solutions on a variety of scenes, as well as some of the applications made possible by rectification.

1. Introduction

The imaging transformation of a perspective camera leads to distortion of a number of geometric scene properties. For instance, length ratios of and parallelism between line segments on a plane in the scene are not preserved in the projected image. As a result, objects appear to grow larger and move faster as they approach the camera centre. Similarly, two objects that are a fixed distance apart in the world will appear to be closer to each other when they are both far away from the camera than when they are both near it. The projective camera thus makes it difficult to characterise objects — in terms of relative sizes, velocities, aspect ratios and so on — and their interactions in a scene.

An automated solution for camera parameters and 3D world coordinates requires two or more camera views. In the visual surveillance domain, where a single static camera is often used, the traditional solution has been to use a precalibrated camera or a special calibration object. With the widespread deployment of low-cost surveillance cameras, such manual techniques are not adequate. Recently, some semi-automatic methods for estimating camera parameters and/or scene geometry using images/video data from a single camera have been proposed [1, 3, 10]. Most of these methods rely on the presence of rich, regular scene geometry to provide calibration information. While it is clearly useful to find the camera parameters, it is difficult to automatically find enough constraints (such as orthogonal or parallel lines) in arbitrary scenes to estimate them all.

In this paper, we propose a set of automatic planar rectification techniques that recover geometric properties for objects moving on a ground plane by tracking the motion of some objects in the scene. We thus restrict our attention to estimation of coordinates on a 2D plane instead of the entire 3D scene. We consider a common situation in which a surveillance camera is mounted well above the ground plane, so that centroids of objects can be approximated as lying on the ground plane. We show that

1. affine rectification[1] of the ground plane (making parallel world lines appear parallel in the rectified image) is possible by tracking two objects that move with constant (but possibly unequal) speeds, if the two objects move along non-parallel linear paths in the scene for some period of time, and

2. metric rectification[1] of the ground plane (making angles in the rectified image equal to angles in the world plane) is possible by tracking two objects moving with constant (but possibly unequal) speeds, if each of the two objects moves along two non-parallel linear paths in the course of its trajectory.

In urban environments, these constraints are often met, since pedestrians and vehicles tend to move in straight lines, and there is typically more than one path of motion in the scene. However, certain other scenes (such as highway or airport scenes) may only contain objects moving along one line. This is a degenerate condition, since 2D information is not available. We show how to detect this case, and recover 1D geometric properties.

After affine rectification of the ground plane, length ratios along parallel lines and arbitrary size ratios can be measured. We discuss a number of applications of these properties, such as finding 'normalised' sizes of objects (and using them to discriminate between vehicles and pedestrians), estimating relative velocities of objects whose trajectories

[1] Affine (metric) rectification means that the rectified image coordinates differ from the world coordinates by an affine (a similarity) transformation.

are parallel (*eg.* cars on a straight road), and achieving Euclidean rectification given a known world distance (such as the length of a particular vehicle).

1.1. Related Work

Some surveillance systems ignore the ground plane rectification problem by restricting themselves to small fields of view, near/mid-field settings or top-down views [5, 9]. These methods are not effective for wide-area surveillance problems such as traffic monitoring or scene-level activity modeling. Existing methods for recovering scene geometry from a single camera view mainly focus on 3D reconstruction and camera calibration, and rely on the presence of rich scene geometry for deriving the necessary constraints. For instance, Criminisi *et al.* [1] recover aspects of 3D affine geometry using hand-labeled scene features such as parallel lines. While this is useful for architectural or forensic purposes, it cannot be employed directly in an automated surveillance system. Jaynes [6] proposes a multi-camera surveillance system in which internal camera parameters are assumed known, and solves for relative pose between each camera and the ground plane using parallel line segments indicated by the user. The need for user interaction is the limiting factor in such approaches.

A few rectification/calibration methods make explicit use of object motion. The calibration technique proposed by Lv *et al.* [10] uses the constraints that pedestrians are essentially perpendicular to the ground plane and that their heights remain constant. This method only works for scenes involving humans walking, and is not applicable, for example, to traffic monitoring. In scenes involving humans and vehicles, a person-detection step is required before calibration. The method also assumes no shadows are present, which, in our experience, is very hard to realise for outdoor scenes. Finally, it cannot work for a top-down view where heights cannot be reliably estimated. In a similar approach, Renno *et al.* [11] propose to use the projected sizes of pedestrians to estimate the vanishing line of the ground plane. In addition to some of the limitations mentioned above, this method is limited by the fact that it uses a linear approximation for the decrease in projected sizes of persons as they move away from the camera. This approximation is valid only for a constrained set of views: in a top-down view, projected sizes of pedestrians actually decrease as they approach the camera centre.

The novelty in our approach lies in the use of object-centroid velocities for deriving the necessary constraints on the image-to-ground homography. Though the observed centroids do not exactly lie on the ground plane, they can be estimated more robustly than points (on the edge of the object silhouette) on the ground plane, as the latter are more susceptible to noise and shadows. Compared to methods based on object sizes, there are fewer parameters to be estimated in our method, and fewer assumptions are needed.

2. Perspective Projection of the Ground Plane

Under perspective projection, the ground plane is mapped to the image plane by a plane projective transformation, or homography. It has been shown that the estimation of such a homography matrix and hence the recovery of geometric properties on the ground plane can be done in a stratified manner [4]. Thus, if points $\mathbf{x}$ on the ground plane, expressed in homogeneous coordinates, are related to image points $\mathbf{x}'$ as $\mathbf{x} = \mathsf{H}\mathbf{x}'$, then the homography matrix H has a unique decomposition into three 3×3 matrices, S, A and P, representing the similarity, affine and pure projective components of the image-to-ground homography respectively:

$$\mathsf{H} = \mathsf{SAP} \tag{1}$$

Our goal is to recover the latter two components and apply them to the image of the scene, to obtain a rectified image that is related to the ground plane by the similarity transformation S (*i.e.* translation, rotation and isotropic scaling). We now discuss the parameters we need to estimate and the world constraints we use for this purpose.

2.1. Constraints on Projection Parameters

The pure-projective component of the image-to-ground homography can be characterised in terms of the homogeneous representation of the vanishing line (also called the line at infinity), $\mathbf{l}_\infty = (l_1, l_2, l_3)^T$, of the ground plane[2][4]:

$$\mathsf{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & l_3 \end{pmatrix} \tag{2}$$

The vanishing line, $\mathbf{l}_\infty$, has two degrees of freedom (since it can be represented as a homogeneous 3-vector), and can be calculated as the line through two or more known vanishing points (or 'points at infinity') of the plane. In the next section, we show that each linear constant-speed path allows us to estimate a vanishing point. Two such non-parallel paths are sufficient (in principle) to estimate $\mathbf{l}_\infty$, and hence also P. Application of the transformation P to the image produces an affine-rectified image which has some nice geometric invariance properties, such as maintaining parallelism between parallel world lines, and keeping area ratios invariant. We demonstrate useful applications based on these recovered properties in Section 5.

Extending affine rectification to metric rectification involves estimation of the affine component A of the image-to-ground homography. It can be shown [4] that this matrix is of the form

$$\mathsf{A} = \begin{pmatrix} \frac{1}{\beta} & -\frac{\alpha}{\beta} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3}$$

[2]The vanishing line of the ground plane in a scene is often called the *horizon line.*

and has two free parameters, α and β, which specify the image of the so-called *circular points* of the plane [4, 8]. The circular points—a pair of complex conjugate points on the line at infinity—are invariant under similarity transformations, but are transformed from metric coordinates $(1, \pm i, 0)^T$ to affine coordinates $(\alpha \mp i\beta, 1, 0)^T$ by the affine transformation A.

Liebowitz and Zisserman [8] have proposed three types of world constraints on the ground plane to estimate α and β: (i) a known angle in the plane, (ii) equality of unknown angles, and (iii) a known length-ratio. We elaborate on the third type of constraint here. Assume that we are given two non-parallel line segments that are represented in image coordinates as the segment joining (x_{11}, y_{11}) and (x_{12}, y_{12}) and the one joining (x_{21}, y_{21}) and (x_{22}, y_{22}). Let the length-ratio of these segments in the world plane be known to be s. It can be shown [7] that, in the 2D complex space with α and β as real and imaginary axes respectively, the point (α, β) lies on a circle with centre $(c_\alpha, 0)$ on the α-axis and radius r:

$$c_\alpha = \frac{\Delta x_1 \Delta y_1 - s^2 \Delta x_2 \Delta y_2}{\Delta y_1^2 - s^2 \Delta y_2^2} \tag{4}$$

$$r = \left| \frac{s(\Delta x_2 \Delta y_1 - \Delta x_1 \Delta y_2)}{\Delta y_1^2 - s^2 \Delta y_2^2} \right| \tag{5}$$

(where $\Delta x_i = x_{i1} - x_{i2}$ and $\Delta y_i = y_{i1} - y_{i2}$). Two such (independent) constraints allow us to solve for α and β. In Section 4, we show that tracking two or more objects in a scene, each of which moves along two non-parallel linear paths with constant speed, allows us to use the above constraint to obtain a metric rectification.

In summary, we propose a unified framework within which both affine and metric rectification can be performed automatically by detecting constant-velocity motion. In the next two sections, we justify the use of constant-velocity constraints, and show how to use them for different types of rectification.

3. Affine Rectification

We use an object tracking system based on Stauffer and Grimson's adaptive backgrounding algorithm [12]. The tracker output consists of pixel values for tracked motion-blobs (regions where motion was detected) and the positions of motion-blob centroids in each frame. Only the centroids (and the timestamps of the corresponding frames) are used for rectification.

3.1. Detecting Constant-Velocity Paths

Object trajectories containing linear constant-speed paths are identified in two steps: identifying all straight-line paths first, and then selecting those which meet the constant-speed requirement. Since lines on the ground plane project to lines in the image under a projective transformation, we select all paths (that is, sequences of centroid positions observed in successive frames) which can be fit by a straight line (up to a least-squares error threshold[3]) and are longer than a certain minimum length in terms of both number of samples observed and number of pixels moved. A selected linear path may correspond to any part of an object's trajectory through the field of view, and need not account for an entire trajectory.

The second step—detecting constant-speed motion—is more challenging, since world speed cannot be measured from projectively distorted images. To get around this problem, we first assume that the linear paths do indeed correspond to constant-speed motion. For each path, we calculate the 1D coordinates, $\mathbf{p}'$, of the observed image centroids (that is, locations of the centroids measured in an arbitrary 1D Euclidean coordinate system fixed to the chosen path-line). We then estimate a 1D homography, represented as a 2×2 matrix G, between the points $\mathbf{p}'$ of each linear path, and an equal number of arbitrarily chosen equidistant 1D positions (corresponding to constant-velocity motion along the world line), $\mathbf{p}$, so as to satisfy

$$\mathbf{p} = \mathsf{G}\mathbf{p}' \tag{6}$$

.

Given a long sequence of centroid positions from a linear path, the matrix G can be estimated by the Direct Linear Transformation algorithm for minimising an algebraic least squares error [4]. We thus obtain one 1D homography for each path. If the reprojection error for a computed homography is large, the corresponding object's motion does not follow the chosen homography model and hence the object clearly does not move with constant speed. If, however, the error is small, the (world) position of the object, $p(t)$, varies as a linear fractional transformation of time t:

$$p(t) = \frac{at + b}{ct + 1} \tag{7}$$

The type of motion then depends on the value of the parameter c. If c is zero, this represents constant-speed motion. On the other hand, if c happens to be non-zero, then Equation 7 represents a type of motion in which neither velocity nor acceleration are constant. There is no way to tell the actual value of c from the observed data. However, since real-world objects are much more likely to move with constant speed than according to Equation 7 over long distances and prolonged intervals of time, we assume that a small reprojection error indicates the presence of a linear, constant-speed path.

3.2. Rectification: Normal Case

Given at least two constant-velocity paths in a scene, we can estimate the imaged line at infinity. The 1D homography calculated for each path in Equation 6 can be used to estimate the position of the imaged vanishing point in that direction, $\mathbf{p_v}'$, by applying the inverse homography to the canonical world vanishing point:

[3]The threshold is proportional to the distance moved, in order to achieve the same fitting accuracy for trajectories of different lengths.

$$\mathbf{p_v}' = \mathsf{G}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (8)$$

If an estimated point $\mathbf{p_v}'$ lies very far away from the endpoints of the detected path in the image, it is not used in the interest of robustness. The line at infinity is then estimated as the best-fitting line in the least-squares sense passing through the vanishing points estimated from the individual constant-velocity paths. The homogeneous vector representing the obtained line at infinity is plugged into Equation 2 to find the pure-projective component, P, of the rectifying transformation. Applying P to the image produces an affine-rectified image. Various applications of this rectified image are discussed in Section 5.

3.3. Rectification: Degenerate Cases

Assuming that a sufficiently large number of objects have been tracked successfully, a degenerate situation might arise if all the linear paths in the scene are parallel in the (Euclidean) world coordinate frame. If this happens, all the vanishing points calculated above (Equation 8) should coincide. Of course, this will not happen in practice due to noise in the data, and the estimated line at infinity will be meaningless. Clearly this situation needs to be detected.

Two types of detected image paths may correspond to a degenerate case:

- paths that are all parallel to each other, or
- paths that converge to a finite point in the image plane.

The first case has already been handled, since the vanishing points in this case will be at infinity (or very far away from the path endpoints) and thus not considered for further processing. The second case is harder, since the detected path-lines in the image may all converge to a finite point, but still not result in degeneracy (if the point of convergence is not the vanishing point of the set of lines). To identify this type of degeneracy, we find the mean vanishing point and calculate the distance of each path-line from it. If all these distances are very small, the imaged path-lines are considered to correspond to parallel paths in the world. If, however, even one of the distances is large, the corresponding path-line is not parallel to the rest of the lines, and hence at least two independent vanishing points exist, which can be used to find the line at infinity. Note that it is important to use tight thresholds while selecting straight lines and constant-velocity paths, to ensure robust estimation of vanishing points.

In a degenerate case of the second type, a single vanishing point, $\mathbf{v_0}'$, for the set of world-parallel paths can be estimated from all the detected image paths, using an iterative optimisation process. In practice, we have found that it is sufficient to take the mean of the estimated vanishing points. With this vanishing point, we have essentially specified one degree of freedom of the line at infinity (since it must pass through this point). We now choose any line passing through the vanishing point $\mathbf{v_0}'$, as long as it does not pass between any pair of observed image paths, and use it as a vanishing line to perform affine rectification. Of course, the rectification will only be valid for measurements along image lines that pass through $\mathbf{v_0}'$. In effect, this gives us a 1D metric rectification. Applications of 1D rectification are discussed in Section 5.

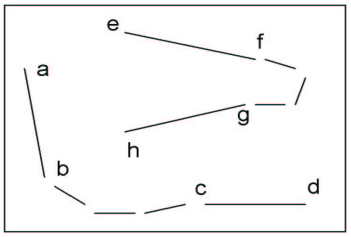

Figure 1: An illustration of two type B trajectories (a–d and e–h) used for metric rectification. Only the detected straight-line paths along the trajectories are shown. Each trajectory (*eg.* a–d) corresponds to an object moving along two constant-velocity non-parallel paths (*i.e.* a–b and c–d). Note that the path followed by the object is not required to be linear or constant-speed in the interval between the two detected constant-velocity paths (*i.e.* in the interval b–c).

4. Metric Rectification

Our procedure for metric rectification of a scene builds on the steps involved in affine rectification. Once again, we use the tracker output to detect constant-velocity paths in the scene. Let us label the corresponding trajectories as Type A. This time, however, we further look for trajectories of objects that contain two paths of constant-velocity motion along non-parallel directions, as illustrated in Figure 1. Let us call these type B trajectories. Clearly, the type B trajectories are a subset of the type A trajectories.

In a real scene, type B trajectories mostly correspond to pedestrian motion. We make one key assumption involving type B trajectories: speeds of the tracked object while moving along the two constant-velocity paths in the trajectory are equal to each other. Making this assumption is crucial in order to use the known length-ratio constraints of Equations 4 and 5 for rectification. While there are clearly cases where this assumption is a bad one—people changing speed as they move from a paved surface to a grass surface, for example—the assumption is often valid in practice.

4.1. Rectification: Normal Case

Given a set of type A and type B trajectories, we use the former to perform affine rectification of the scene, as discussed in Section 3. All further measurements are made on the affine-rectified image obtained. For each type B trajectory, $\mathcal{B}_i$, $1 < i < N_{\mathcal{B}}$, a length-ratio s_i is estimated as the ratio of time-intervals spent moving along the two constant-velocity paths in the trajectory. From Equations 4 and 5, each of these s_i leads to a circular constraint C_i on parameters α and β of the affine component, A of the image-to-ground homography. The solution for these parameters is

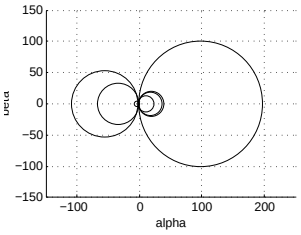

Figure 2: An example set of circular constraints on the parameters α and β of the affine component of the image-to-ground homography.

given by intersection of the $N_{\mathcal{B}}$ circles, $\mathbf{c}_i$ (as illustrated in Figure 2). Of course, more than two circles will not intersect in two unique points because of noise in the estimates of centres and radii. To get around this problem, we simultaneously estimate the two intersection points and $N_{\mathcal{B}}$ circles passing through them by iteratively minimising the errors in centres and radii (details omitted due to lack of space).

4.2. Rectification: Degenerate Cases

A degenerate case for metric rectification arises whenever the constraints provided by the Type B trajectories are not independent. In the problem formulation given above, this can happen if the perpendicular-bisector of the angle between the pair of constant-velocity directions from one trajectory is either parallel or perpendicular to the perpendicular-bisector calculated similarly from the other trajectory. In the absence of noise, dependent constraints lead to identical circles. Working with real data, we have found that simply using a threshold on the minimum spread in values of radii and centre coordinates (along the α axis), we can sufficiently well identify degenerate cases. It should be noted that, even in a non-degenerate case, the estimated point of intersection may be highly sensitive to measurement errors if the radii of all the circles used for estimation are large. The accuracy of estimation can thus be judged by the number of circles whose radii are less than a specified threshold.

5. Applications of rectification

Given an affine rectification, relative areas can be measured for any two vehicles or persons at arbitrary positions in the image. However, velocities can only be compared along world-parallel lines. With metric rectification, velocities along arbitrary directions can be compared.

Let us call the projected image area of an object in an affine-rectified image its 'normalised' size. The normalised size of an object in a scene is related to its 'actual' 2D size (as observed when looking top-down) by a fixed scale factor. Normalised sizes can be used to easily classify objects in a scene, as demonstrated in Section 6.

It is important to appreciate the approximation involved in the above application: real-world objects are non-planar. This means that the normalised sizes of objects correspond to their projection on to a scene plane (when viewed from the camera direction). While the planar approximation is clearly valid if the camera is situated well above the ground plane, it should be noted that the scene plane in question actually corresponds to a plane passing though the object centroids, since these are the points used for rectification. Projection on to this 'centroid-plane', which is situated at about half the height of an average pedestrian, causes less distortion than projection on to the actual ground plane.

In a degenerate 1D case, the only meaningful metric properties that can be measured are along a direction parallel to the bundle of lines used for rectification[4]. Thus, relative lengths and speeds of vehicles moving in parallel lanes, as well as relative speeds of vehicles on the road and pedestrians on the footpath, may be estimated. Knowing the true length of a particular vehicle, the 1D scale factor can be fixed, so that actual speeds of vehicles can be measured. This can be used, for instance, to automatically detect and track a speeding car on a highway. In fact, if a vehicle of known length moves along two non-parallel directions in a scene, an affine rectification of the scene can be upgraded to metric. Even if no single vehicle can be automatically identified, a prior on the probability distribution of lengths of vehicles in typical road scenes can be used to find the best scale factor in a statistical sense.

6. Experiments

As mentioned earlier, we use Stauffer and Grimson's tracking algorithm [12] to obtain object trajectories. The tracker frame-rate averages 8 frames a second at a resolution of 320x240 pixels. Objects occupying more than 25% of the frame area, or objects visible for less than a second, are automatically filtered out before further processing. The search for constant-velocity paths is run on the tracking data till at least three vanishing points along non-parallel world lines are found[4]. Image paths for which the root-mean-square line-fitting error is less than 2% of the path-length are considered straight. Further, straight-line paths for which the reprojection error of the 1D homography is less than 2% of the path-length are considered constant-velocity paths. Constant-velocity paths that are at least two-fifths of the frame height and 25 frames or longer are selected for estimating the vanishing line. Affine-rectified images for a set of scenes, along with the corresponding vanishing points and lines, are shown in Figure 3. Note that the technique works well for different types of scenes, as can be judged from the recovered parallelism of roads and footpaths. To judge the quality of rectification, we calculated the standard deviation in normalised sizes of objects, which turned out to be less than 3% of the normalised size on average. Size constancy is illustrated in Figure 6.

The conditions required for metric rectification of a scene are slightly more complex, and thus more scene dependent. Figure 4 shows an example of automatic metric rectification. Ground truth data, in the form of an archi-

[4]The procedure described in Section 3.3 is used to detect parallelism.

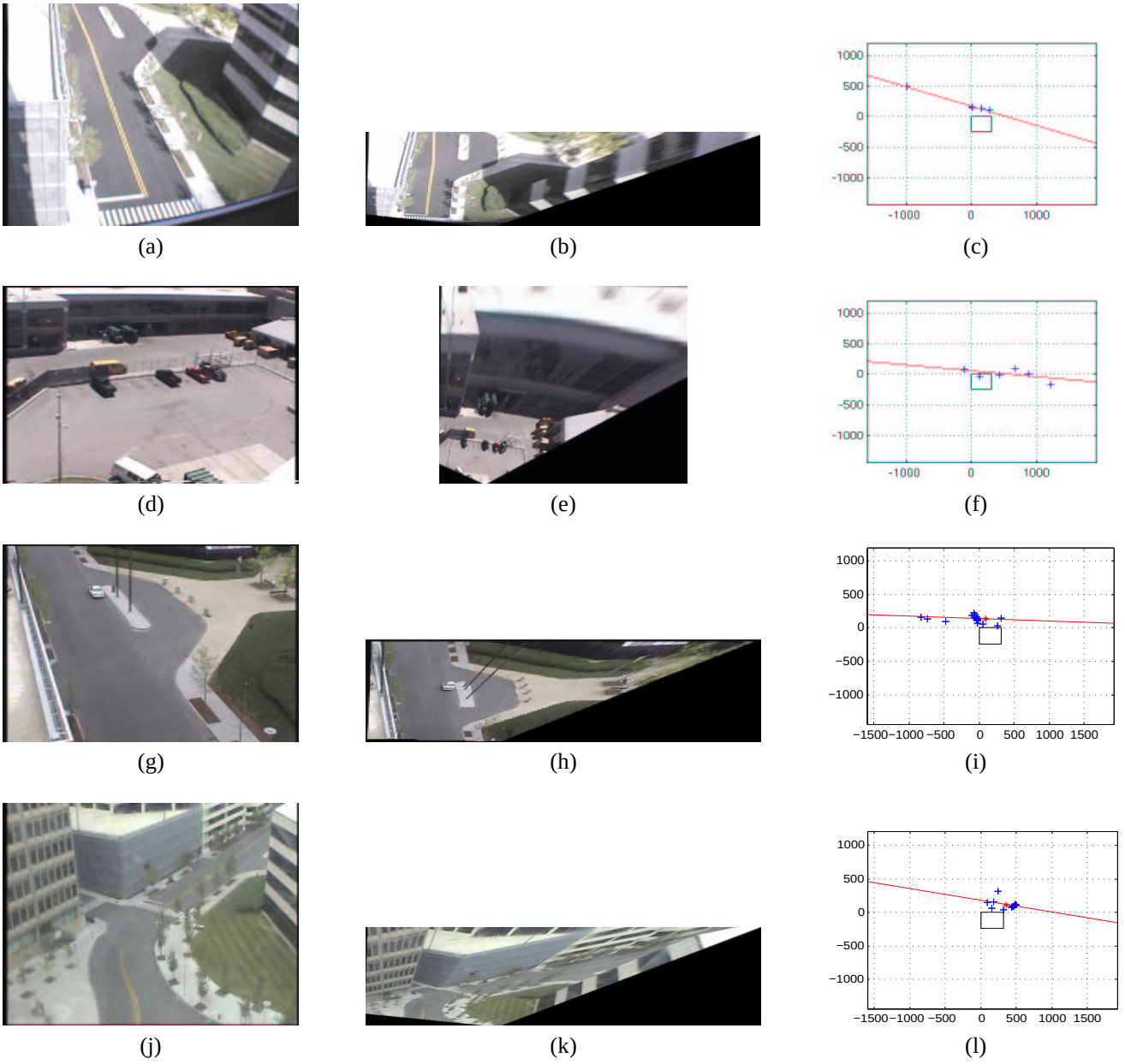

Figure 3: Examples of affine rectification: original scenes (a,d,g,j) and corresponding affine-rectified images (b,e,h,k) and affine geometries (c,f,i,l). The black rectangle in the centre of each geometry plot marks the image region; the blue crosses are estimated vanishing points along constant-velocity paths, and the red line is the estimated vanishing line; axis coordinates are in pixels. For display purposes, some images have been cropped to remove parts that are not on the ground plane. Notice how the road edges are parallel after rectification. (The lane-dividing line in scene (a) is not a straight line in the world.)

Figure 4: Metric rectification. (1) Original image, (2) metric-rectified image and (3) architectural plan of the scene. Compared to the corresponding affine-rectified image (Figure 3(h)), note the difference in aspect ratio of the traffic island.

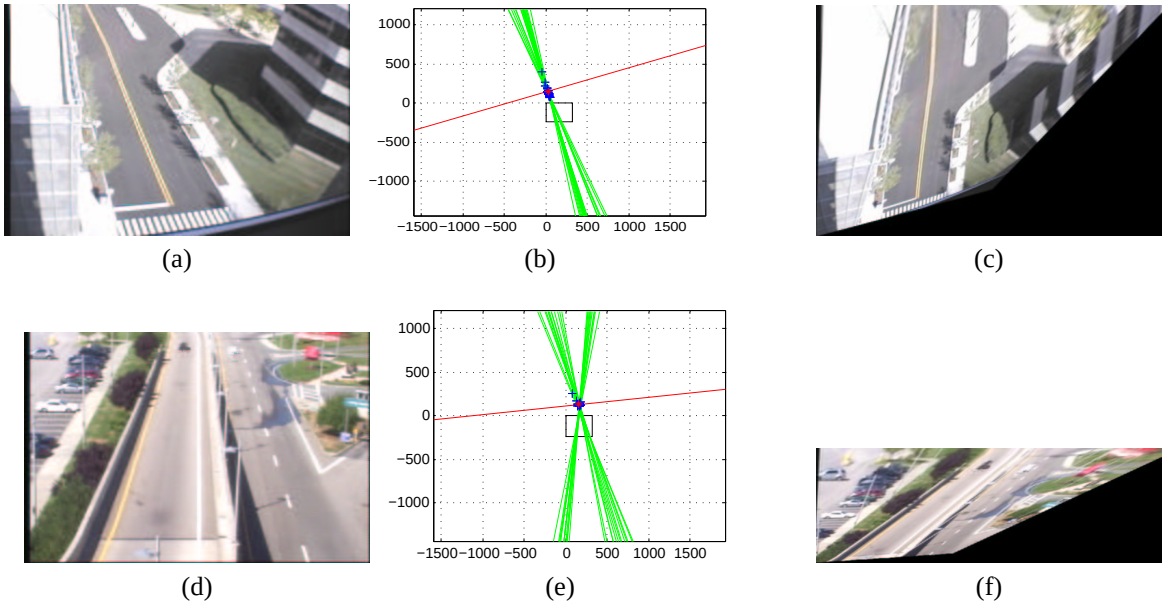

Figure 5: Scenes with degenerate object motion (a,d), the corresponding affine geometry (b,e) and 1D rectification (c,f). The black rectangles in the centres of the plots (b,d) mark the image regions. Straight lines along which constant-velocity paths were detected are in green, and blue crosses are corresponding vanishing points. The red dot is the mean vanishing point: note how all the straight lines pass very close to it. The red line is an arbitrary line passing through the mean vanishing point, as described in Section 3.3. Only geometric properties parallel to the roads can be measured correctly in the 1D rectified images. Axis coordinates are in pixels. (The lane-dividing line in scene (a) is not a straight line in the world.)

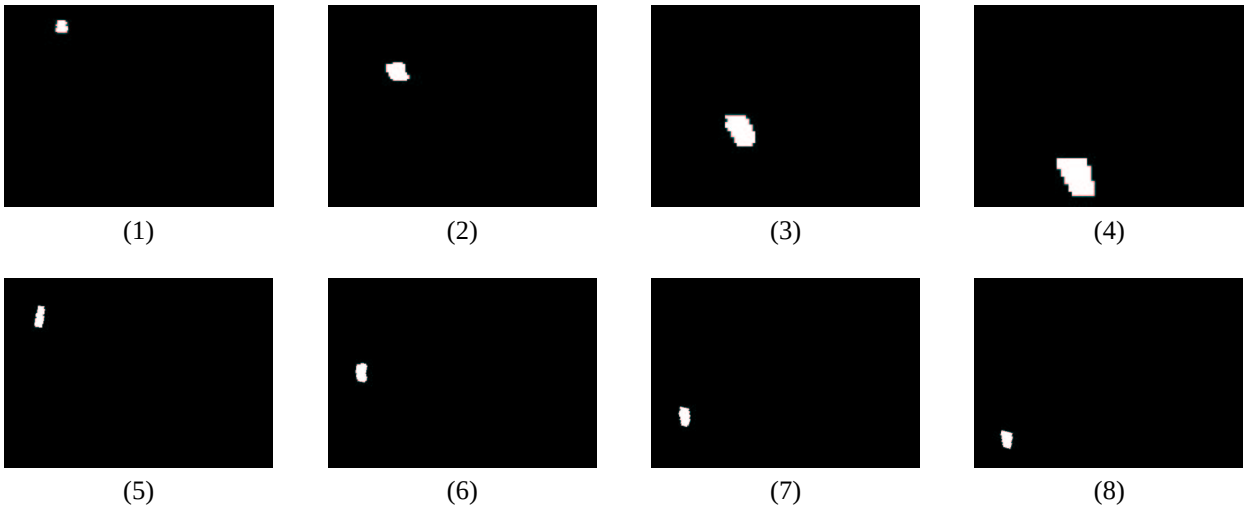

Figure 6: Original projected image silhouettes of a vehicle (1–4), and projected silhouettes after affine rectification (5–8) in scene (a), Figure 3.

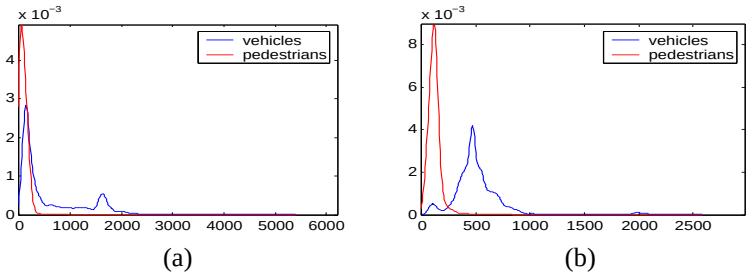

Figure 7: Class-conditional projected-size densities of vehicles and pedestrians in scene (a) of Figure 3, (a) before and (b) after affine rectification. Horizontal axis coordinates are in pixels.

tectural plan of the scene, are shown alongside the rectified image for comparison. Since most of the applications we propose work for affine (or even 1D) rectification, we mainly regard metric rectification as a source of extra information whenever it is achievable. In the extreme, our metric rectification technique can be used as a fast method of manual rectification, by having a few people walk at (almost) constant speed in different directions in the scene.

If the required constraints for affine rectification are not found even after observing 100 constant-velocity paths in a scene, a degenerate condition probably exists. Detection of degenerate 1D-motion cases is illustrated in Figure 5. The first example corresponds to a manually filtered data set for the scene in Figure 3(a). The only objects in the filtered data set are vehicles moving on the road. To detect degenerate cases, each path-line passing within 30 pixels of the mean vanishing point is considered to belong to a bundle of world-parallel lines, as explained in Section 3.3. The second example illustrates the same method for a scene in which there are two spatially separate groups of lines that form a bundle of world-parallel lines. Note that the parallelism of the edges of the road and the footpath are recovered in the 1D rectified images.

To show that the size normalisation produced by affine rectification can simplify vehicle/pedestrian classification, we estimated class-conditional densities for the two classes, vehicles and pedestrians, in scene (a), Figure 3, using a non-parametric density estimator (Parzen windows [2]). As seen from the estimated conditional densities (Figure 7), the overlap between the two distributions (and hence the Bayes error) is greatly reduced if normalised sizes are used. We also trained a linear SVM classifier [2] on 700 vehicle and person images, and tested on 5500 images. When trained on original image sizes, the test errors were 41% and 7% for vehicles and persons respectively. By training on normalised sizes instead, these errors were reduced to 7% and 4% respectively. To illustrate Euclidean calibration in this scene, we assumed an average vehicle length of 14 feet and estimated the average speeds of vehicles and persons in this scene as 11 and 2 miles per hour respectively.

7. Conclusions

A novel technique for recovering affine and metric properties of a ground plane, based on tracking moving objects, has been presented. The method requires no knowledge of camera parameters or the types of objects in the scene. The only significant approximation made is that centroids of the imaged objects lie on a single world plane. The key assumptions are that objects do not move in a manner such that their position must be expressed as a linear fractional transform of time (as in Equation 7), and that pedestrians do not change pace by a fixed factor when moving from one part of a scene to another. The method has been found to be applicable to many outdoor scenes, as constant-velocity motion is quite common in the real world. A 1D rectification scheme has been proposed for situations where 2D rectification is not possible.

We are working on developing further interesting applications based on size normalisation of tracked objects, such as relative calibration of a multi-camera system from statistical characterisations of object-size distributions.

Acknowledgments

Tracking data were provided by Chris Stauffer. Biswajit would like to thank Gerald Dalley and Peter Sturm for useful discussion on the feasibility of the proposed approach, and Andrew Zisserman for advice on the problem of robustly estimating intersection of circles for metric rectification.

References

[1] Criminisi, A., Reid, I. and Zisserman, A., "Single View Metrology," *Proc. 7th Intl. Conf. on Computer Vision (ICCV)*, 1999.

[2] Duda, R.O., Hart, P.E. and Stork, D.G., *Pattern Classification*, 2nd ed., John Wiley and Sons, Inc., 2001.

[3] Gurdjos, P., Payrissat, R., "About Conditions for Recovering the Metric Structures of Perpendicular Planes from the Single Ground Plane to Image Homography," *15th Intl. Conf. on Pattern Recognition (ICPR)*, 2000.

[4] Hartley, R.I. and Zisserman, A., *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[5] Javed, O. and Shah, M., "Tracking and Object Classification for Automated Surveillance," *Proc. European Conf. on Computer Vision (ECCV)*, 2002.

[6] Jaynes, C., "Multi-view Calibration from Planar Motion for Video Surveillance," *Proc. Intl. Workshop on Video Surveillance*, 1999.

[7] Liebowitz, D., "Camera Calibration and Reconstruction of Geometry from Images," D. Phil. thesis, University of Oxford, 2001.

[8] Liebowitz, D. and Zisserman, A., "Metric Rectification for Perspective Images of Planes," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 1998.

[9] Lipton, A.J., Fujiyoshi, H. and Patil, R.S., "Moving target classification and tracking from real-time video," *Proc. IEEE Workshop on Applications of Computer Vision (WACV)*, 1998.

[10] Lv, F., Zhao, T. and Nevatia, R., "Self-Calibration of a camera from video of a walking human," *Proc. Intl. Conf. on Pattern Recognition (ICPR)*, 2001.

[11] Renno, J., Orwell, J. and Jones, G.A., "Learning Surveillance Tracking Models for the Self-Calibrated Ground Plane," *Proc. British Machine Vision Conf. (BMVC)*, 2002.

[12] Stauffer, C. and Grimson, E., "Learning patterns of activity using real-time tracking," *Proc. IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, 2000.