View Independent Object Classification Based on Automated Ground Plane Rectification for Traffic Scene Surveillance

Zhaoxiang Zhang, Min Li, Kaiqi Huang, Tieniu Tan

View Independent Object Classification Based on Automated Ground Plane Rectification for Traffic Scene Surveillance

Zhaoxiang Zhang, Min Li, Kaiqi Huang and Tieniu Tan
National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences
{zxzhang, mli, kqhuang, tnt}@nlpr.ia.ac.cn

Abstract

We address the problem of view independent object classification. Our aim is to classify moving objects of traffic scene surveillance videos into pedestrians, bicycles and vehicles. However, this problem is very challenging due to large object appearance variance, low resolution videos and limited object size. Especially, perspective distortion of surveillance cameras makes most 2D object features like size and speed related to view angles and not suitable for object classification. In this paper, we adopt the common constraint that most objects of interest in traffic scenes are moving on the ground plane. Firstly, we realize the ground plane rectification based on appearance and motion information of moving objects, which can be applied for normalization of 2D object features. An online learning framework is then described to achieve automatic object classification based on rectified 2D object features. Experimental results demonstrate the effectiveness, efficiency and robustness of the proposed method.

1. Introduction

Automatic object classification in videos is an important issue in computer vision and visual surveillance with great potential for real applications. With objects type information known, more specific and accurate methods can be developed to monitor high level actions of moving objects. Especially for traffic scene surveillance, classification of moving objects into predefined categories allows the operator to program the monitoring system by specifying events of interest for different objects types like 'alarming when a pedestrian is coming into a forbidden area' or 'alarming when a vehicle is running in a reverse direction', which is very common in intelligent visual surveillance.

Due to its importance, much work has been done on automatic object classification. In [2, 5, 7, 17], shape features like area, compactness, bounding box aspect ratio and motion features like speed and motion directions are extracted for training and classification. However, these shape and motion features are based on image plane so that they cannot avoid perspective distortion, which is much more significant in far-field traffic scene surveillance videos. For example, nearby objects in images appear to be larger and move faster than those far away. Therefore, simply using these 2D object features for classification is not suitable and limits the accuracy of object classification. In [15], video scenes are divided into subregions to be treated respectively to decrease the effect of perspective distortion. In [6], a series of algorithms are described to demonstrate the effectiveness of local features for object detection and classification. However, most of the algorithms are time-consuming and not applicable to low resolution videos. In [8], SVM is applied with Histogram of Orientated Gradient features for classification. Viola et al [14] give us a good framework for automatic feature selection and object classification by Boosting. However, this framework needs to collect large samples of training data in all kinds of conditions and label all of them manually.

As we have described above, 2D object features suffer from perspective distortion, which make original 2D feature not applicable to object classification. How to conquer the effect of perspective distortion is the key to use 2D features for efficient object classification. The common solution to perspective distortion is to use a pre-calibrated camera. However, manual calibration always need a wide site survey of surveillance scenes [12] and limits the practicality of classification algorithms to different scenes. Various approaches [3, 9, 11] are proposed for auto-calibration from inherent scene structures or accurate pedestrian detection. However, inherent structures are related to surveillance scenes and precise pedestrian detection is very challenging in low resolution surveillance videos.

In fact, there is a common constraint that most objects of interest in traffic scene surveillance videos are moving on or near the ground plane. We can deal with perspective distor-

tion of cameras based on homography between the ground plane and image plane. Stauffer et al [13] achieved normalization of tracking data on an inaccurate linear assumption. Bose et al [1] achieved metric rectification of the ground plane by extracting a series of moving objects along linear path with constant speed. However, the conditions are rigor to be satisfied.

In this paper, we solve perspective distortion of surveillance cameras based on a robust automated ground plane rectification which has already been proposed in detail in [16]. Firstly, both affine rectification and metric rectification are achieved based on appearance and motion information of moving objects. These rectifications can be effectively used for normalization of 2D object features. A novel online learning framework is then applied to make use of rectified 2D object features for classification. Experimental results demonstrate the effectiveness, efficiency and robustness of our approach.

The remainder of the paper is organized as follows. In Section 2, we briefly introduce the principle of the ground plane rectification. In Section 3, we introduce the method of affine rectification. Details of this method in the case of straight roadways was proposed in [16]. Here we extend the method to the cases of non-straight roadways. Then the principle of metric rectification is introduced in Section 4. In Section 5, we propose our online learning framework for classification. Experimental results and analysis are presented in Section 6. Finally, we draw our conclusions in Section 7.

2. Ground Plane Rectification

Under perspective projection, the ground plane is mapped to image plane by homography. Points on image plane, $\mathbf{x}$, are related to points on the world plane, $\mathbf{x}'$, as $\mathbf{x}' = \mathbf{H}\mathbf{x}$. As has been described in [12], the homography matrix $\mathbf{H}$ can be decomposed uniquely into three matrices, $\mathbf{S}$, $\mathbf{A}$, $\mathbf{P}$, representing the similarity, affine and pure-projective components of homography respectively:

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

The similarity component $\mathbf{S}$ is a similarity transformation which has no relation to the affine and metric rectification.

The pure-projective component is characterized by a vanishing line $l_\infty = (l_1, l_2, l_3)^T$ of the ground plane, which has the as:

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

Recovery of the pure-projective component $\mathbf{P}$ achieves affine rectification of the ground plane.

Extending affine rectification to metric rectification involves estimation of the affine component $\mathbf{A}$ with the form:

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

This matrix has two degrees of freedom represented by α and β, which specify the image of the circular points [12]. The circular points are invariant under similarity transformations, but are transformed from metric coordinates $(1, \pm i, 0)^T$ to affine coordinate $(\alpha \pm i\beta, 1, 0)^T$ by the affine transformation $\mathbf{A}$.

The above part in this section is referenced to [16] and the approach to realize affine and metric rectification respectively from moving objects in traffic scene surveillance videos are described as follows.

3. Affine Rectification

Affine rectification of the ground plane requires identification of the vanishing line l_∞ of the ground plane, which can be determined by two horizonal vanishing points. In this section, we propose our approach to recover two vanishing points of the ground plane based on moving vehicles in traffic scene surveillance videos.

3.1 Coarse Moving Vehicle Detection

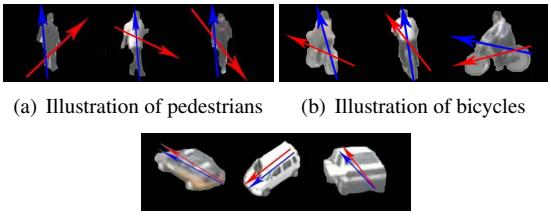
Moving objects in traffic scenes can be detected accurately with shadows removed by improved GMM [10], but we need to distinguish vehicles from pedestrians further more. The difference of the following two directions are taken as a distinctive feature for coarse vehicle detection. The first direction is the velocity direction of objects in videos, which can be calculated due to position change of unit time. The second one is the main axis direction θ, which can be estimated from moment analysis of silhouette:

$$\theta = arctan(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}) \tag{4}$$

Here, μ_{pq} is the central moment of order (p, q). It is evident that the angle difference is very small for moving vehicles while it is significant for pedestrians and bicycles as illustrated in Figure 1. Instead of K-Mean clustering, we adopt a more reliable strategy that only those objects with angle difference less than $\theta_T = 5^\circ$ are labeled as vehicles with all of the rest discarded. The latter estimation of vanishing points benefits from this strict detection strategy.

3.2 Linear Equation Estimation

In most view angles of surveillance scenes, vehicles in videos are rich in line segments along two orientations cor-



(a) Illustration of pedestrians (b) Illustration of bicycles

(c) Illustration of vehicles

Figure 1. Direction difference (Red arrowhead stands for velocity direction; blue arrowhead stands for main axis direction)

responding to the symmetrical axis direction and its perpendicular direction. We make use of image gradient to extract these two accurate line equations for every vehicle detected from videos. As shown in Figure 2, these two orientations are extracted by two stages of Histogram of Orientation Gradient (HOG) and the respective line equations are determined by correlation to image data. Motion direction can help us to distinguish these two lines. The one with orientation close to motion direction corresponds to the symmetrical axis direction while the other one corresponds to its perpendicular direction.

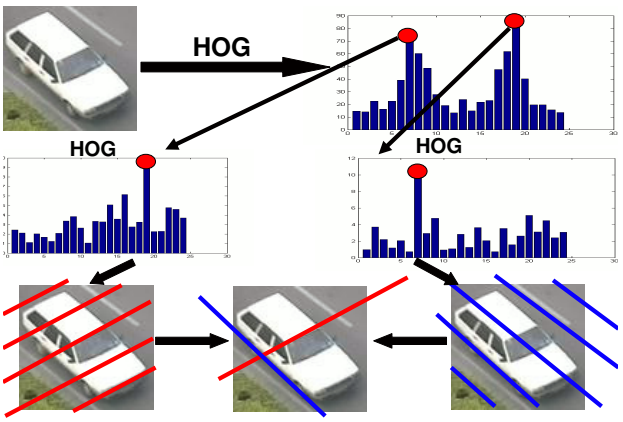
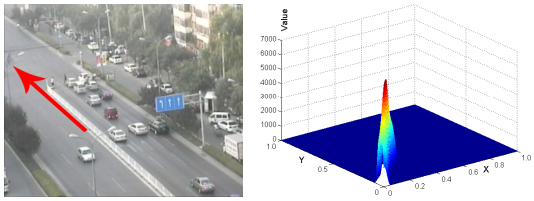


Figure 2. Estimation of line equations from vehicles (cited from [16])

3.3 Intersection Estimation

It is common that most vehicles in traffic scenes are moving along roadways which are mostly straight or contain a series of approximately straight line segments in the view field. Here, we assume that there is only one straight roadway in the view field. So, symmetrical axis of most vehicles should be parallel to each other in 3D world. Due to image projection, they are no longer parallel but intersect to the same point called horizonal vanishing point on image plane. The perpendicular direction is of the same case. We make use of voting strategy to estimate these two horizonal vanishing points. For every line l extracted from vehicles, each point $s(x, y)$ lying on l generates a Gaussian impulse in voting space with (x, y) as its center. With time accumulated, a voting surface is generated and the position of its global extreme corresponds to the estimated intersection point. One example of voting space corresponding to the roadway direction is shown in Fig. 6.



(a) Traffic scene (b) Voting surface

Figure 3. Illustration of estimating vanishing points (cited from [16])

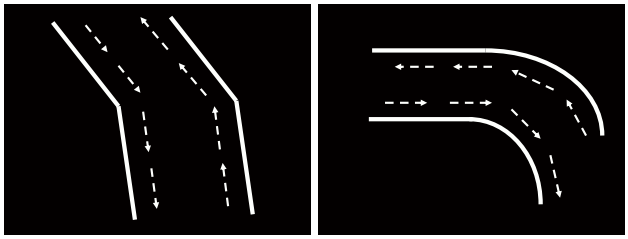
3.4 Special Cases of Roadway Layout

The above part in this section is referenced to [16]. However, the solutions have only discussed to the assumption of only one straight roadway in view field. However, this assumption is not always true in reality. There may be more than one roadway in the view filed, like a crossroad. The roadway may be not straight at all, like a bend. In these cases, the method described above cannot applied to estimate the two horizonal vanishing points. Fortunately, the variance of roadway layouts in reality can be seen as combinations of several primitive layouts which can be solved for horizonal vanishing points estimation and are discussed respectively as follows.

3.4.1 Straight Segments

In reality, it is very common that the whole roadway is not straight, but is contains series of straight segments. Fig. 4(a) shows the case of one inflexion which divides the whole roadway into two straight segments. Fig. 4(b) shows a roadway which is consist a straight segment and a bend. The roadway in these cases contains at least one straight segment, which can be applied for horizonal vanishing points estimation.

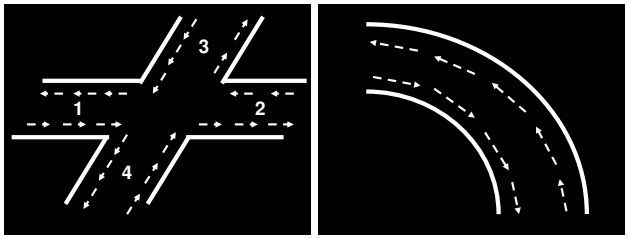
The detection of straight segments can be simply realized by motion information. With objects extracted by mo-



(a) Case 1 (b) Case 2

Figure 4. Straight segment cases

tion detection and classified as vehicles, conventional tracking can help us to monitor the change trend of velocity directions. As we know, the velocity direction of a vehicle changes little in a straight segment but bound at the inflexion or bend part. In this way, we can detect all straight segments from scenes and each straight segment can estimate two orthogonal horizonal vanishing points.



(a) The case of crossroad (b) The case of bend

Figure 5. Crossroad and bend

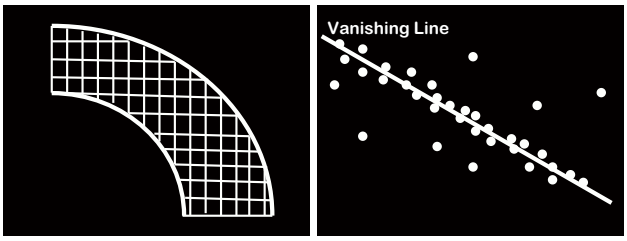
3.4.2 Crossroad

The case of crossroad is more complicated as illustrated in Fig. 5(a). The activities of vehicles in the crossroad contains: running ahead, turning left, turning right and turning around. A trusty strategy is similar to the straight segment case. For the crossroad shown in Fig. 5(a), we can detect four straight segments by motion direction information. For every straight segment, we can estimate a couple of two orthogonal horizonal vanishing points with the strategy of the conventional case. Evidently, the couples estimated from Segment 1 and Segment 2 should be approximate to each other while those of Segment 3 and Segment 4 should be approximate to each other.

In fact, we can still use the strategy of conventional case for vanishing points estimation for crossroad. In this case, there will be two evident peaks in the voting spaces for horizonal vanishing points estimation. One peak corresponds to the direction of Segment 1 and 2 while the other peak corresponds to the direction of Segment 3 and 4. Traffic flow in crossroad are always attempered regularly to avoid accidents. As a result, the two peaks should be of different height so that they can be distinguished from each other. This strategy can also recover two groups of orthogonal horizonal vanishing points.

3.4.3 Bend

Now we discuss the case of a bend as shown in Fig. 5(b). It is the most complicated case because we cannot detect even one straight segment from this roadway. Instead, we can detect the roadway region by vehicle motion information and divide the whole roadway randomly into many small pieces as shown in Fig. 6(a). If the vehicle with its centroid within the piece, we confirm that the vehicle is passing by the piece. We assume that most vehicles passing by one pieces should move in the same direction. As a result, we can estimate a couple of two horizonal orthogonal vanishing points for every piece by the strategy of the conventional case based on motion and appearance information of vehicles passing by the very piece. Since we divide the whole roadway into so many small pieces, those pieces with enough vehicles passing by can supply us a large number of couples of horizonal vanishing points. Since the assumption of vehicles moving in the same direction in one piece is not very accurate, we should select the most accurate couples from the large set.



(a) Division of pieces (b) Estimation of horizonal vanishing line

Figure 6. Illustration of roadway division and estimation of vanishing line

As we know, all horizonal vanishing points should lie on the horizonal vanishing line. For all the horizonal vanishing points estimated from every pieces, Hough Transform can be applied to estimate this vanishing line, which is illustrated in Fig. 6(b). Those couples with the smallest sum of distances to the estimated vanishing line are taken as effective horizonal orthogonal vanishing points.

For one straight roadway, we can estimate exclusive couple of orthogonal horizonal vanishing points from appearance and motion information of moving objects in videos. Since other cases of roadway layouts can be seen as combinations of the above three primitive layouts, we can estimate

several couples of horizonal vanishing points and select the most accurate couple from them. The two horizonal vanishing points determine the horizonal vanishing line, which achieves the affine rectification of the ground plane.

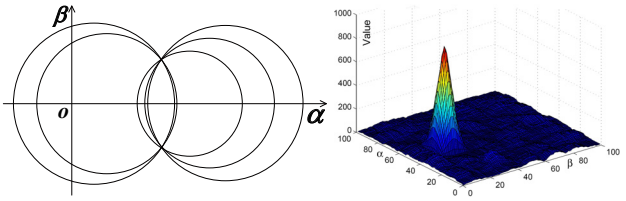
4. Metric Rectification

As described in [4], each known angle θ on the world plane between line l_a and line l_b on image plane gives a constraint of (α, β) to lie on a circle with center (c_α, c_β) and radius r:

$$(c_\alpha, c_\beta) = (\frac{(a+b)}{2}, \frac{(a-b)}{2} cot\theta), \quad r = |\frac{(a-b)}{2sin(\theta)}| \quad (5)$$

where $a = -l_{a2}/l_{a1}$ and $b = -l_{b2}/l_{b1}$ are the line directions.

As described above, each detected vehicle gives two perpendicular directions on the world plane to determine a circle about (α, β). Since there are redundant detected vehicles from videos, we can determine (α, β) as the intersection of a large set of estimated circles as shown in Fig. 7(a).



(a) Illustration of circles (b) Voting surface

Figure 7. Illustration of estimating intersections of circles (cited from [16])

Due to symmetric property, we only focus on the intersection above the α axis. Every two circles can determine the intersection conveniently by differential of these two circle equations. For N circles, we can obtain $N(N-1)/2$ candidate points and (α, β) is determined simply based on Gaussian voting strategy described above as shown in Fig. 7(b). With (α, β) estimated, we can calculate the affine matrix **A** so that the metric rectification is realized. The above part in this section is referenced to [16].

5. Object Classification

In this section, we describe our approach for view independent object classification. The flowchart of the method is shown in Figure 8.

In our framework, we mainly use the following five shape and motion features:

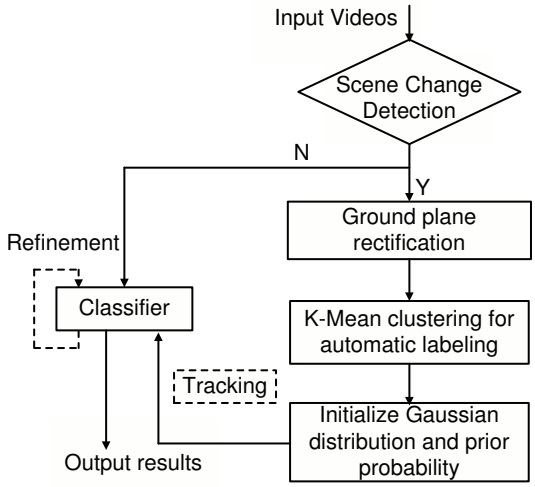


Figure 8. Classification flowchart

- $size$: size of objects in pixels
- $velocity$: time derivative of centroid of the object
- $compactness$: equals to $\frac{area}{perimeter^2}$
- $size'$: time derivative of $size$
- $angle$: angle between motion direction and direction of major axis of the silhouette

Most of these features are efficient for classification in near field videos, but not available in far field videos due to significant perspective distortion. Fortunately, the ground plane rectification enables normalization of these features to be independent of view angles, which are denoted as ($size_{\mathrm{R}}$, $velocity_{\mathrm{R}}$, $compactness_{\mathrm{R}}$, $size'_{\mathrm{R}}$, $angle_{\mathrm{R}}$).

The online classification framework is composed of three periods. The first period achieves the ground plane rectification based on motion and shape information of coarsely detected vehicles. The second period achieves automatic object type labeling based on unsupervised K-Mean clustering. The third period achieves online object classification and refinement of classifiers.

5.1 Automatic Labeling

In order to classify moving objects automatically without supervised learning, we adopt K-Mean clustering and decision level fusion for automatic labeling. We use the three features ($compactness_{\mathrm{R}}$, $size'_{\mathrm{R}}$, $angle$) for K-Mean clustering and automatic labeling. After videos processed frame by frame for a period of time, K-Mean clustering is adopted to establish 3 clusters. Each cluster corresponds to each category, respectively. The decision level fusion based on the following three intuitive rules is adopted to establish the correspondences: (1) $compactness_{\mathrm{R}}$ has the advantages of distinguishing vehicles from pedestrians and bicycles. (2) $area'_{\mathrm{R}}$ has the advantages of distinguishing

pedestrians from vehicles and bicycles. (3)$angle$ has the advantages of classifying pedestrians and vehicles. Using voting strategy, we can conveniently achieve automatic labeling.

5.2 Bayesian Classification

After rectification, the 2D features are independent of view angle changes. Here we make an assumption that $v = (area_{\mathrm{R}}, speed_{\mathrm{R}}, compactness_{\mathrm{R}})$ of every category satisfies a multivariate Gaussian distribution. The assumption will be tested in Section 6 and the distributions are denoted as:

$$P_i(v) = \eta(v, \mu_i, \Sigma_i) \quad i = 1, 2, 3 \tag{6}$$

Using Bayesian rules, we obtain the derivation as follows:

$$P(category = i|v) \propto P_i(v) \cdot p_i \quad i = 1, 2, 3 \tag{7}$$

where $P(category = i|v)$ and p_i are posterior and prior probability of each category, respectively. The prior probability of each category is initialized by the number of individuals belonging to each cluster after automatic labeling period and the Gaussian distribution is estimated from each cluster in the following way:

$$\hat{\mu}_i = \frac{1}{N}\sum_{r=1}^{N} v_{i,r} \quad i = 1, 2, 3 \tag{8}$$

$$\hat{\sigma}^2_{ij} = \frac{1}{N}\sum_{r=1}^{N}(v_{i,r} - \hat{\mu}_i)(v_{j,r} - \hat{\mu}_j) \quad i, j = 1, 2, 3 \tag{9}$$

The category is determined by the posterior probability and the classifier is refined at the same time to be robust to condition changes:

$$p_{k,new} = (1 - \beta)p_{k,old} + \beta(M_{k,t}) \quad k = 1, 2, 3 \tag{10}$$

$$\hat{\mu}_{new} = (1 - \gamma)\hat{\mu}_{old} + \gamma v_t \tag{11}$$

$$\hat{\sigma}^2_{i,j,new} = (1 - \gamma)\hat{\sigma}^2_{i,j,old} + \gamma(v_{i,t} - \hat{\mu}_t)(v_{j,t} - \hat{\mu}_t) \tag{12}$$

where β and γ are the refinement rate. $M(k, t)$ is 1 if v_t is classified as the category k and 0 otherwise. The prior probability is renormalized after that. In every frame, there is posterior probability output for every moving objects. We can determine the category using the sum of posterior probability of tracked frames to improve robustness of classification.

5.3 Discussion

Most scene changes in video surveillance are abrupt transitions caused by zooming or moving of cameras rather than gradual transitions. We can simply detect scene changes when

$$\Sigma_{x,y}(|B_t(x, y) - B_{t-1}(x, y)|) > T \tag{13}$$

where T is a threshold and B_t and B_{t-1} are recovered background of the current and previous frames, respectively. As we use reflectance component for background modeling, the detection is robust to fast illumination changes. When scene changes are detected, we can re-rectify the ground plane and initialize the classifier.

Further more, the classification results can feedback to the ground plane rectification. The objects which are classified as vehicles can substitute coarse vehicle detection and contribute the the affine and metric rectification of the ground plane.

6. Experimental Results and Analysis

Numerous experiments are conducted and experimental results are presented in this section to demonstrate the performance of the proposed approach. All experiments are conducted on a computer of P4 3.0 CPU and 512M DDR.

6.1 Illustration of Appearance Rectification

As we have described before, object appearance has significant distortion due to camera projection. The ground plane rectification enables normalization of appearance. Here, we take object silhouette as an example to illustrate the performance of appearance rectification. Experiments are conducted to a vehicle moving across a far-field traffic scene. The original projected silhouettes and corresponding rectified silhouettes are shown in Figure 9. As we can see, after rectification, the object silhouette is approximately invariant in far-field scenes, which illustrates the effectiveness of appearance rectification. This part is referenced to [16].

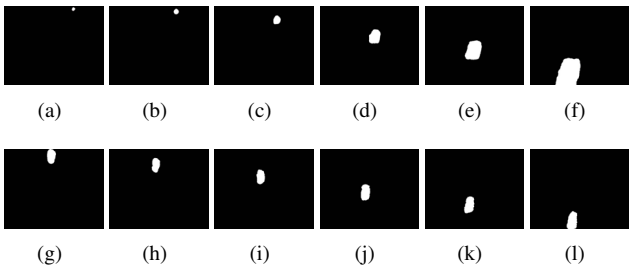
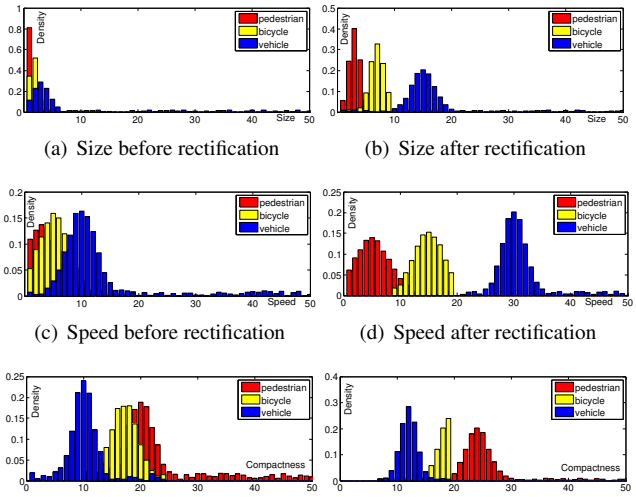


Figure 9. Rectification of silhouette for a moving vehicle ((a)-(f) are original silhouettes; (g)-(l) are rectified silhouette)

6.2 Illustration of Feature Rectification

The ground plane rectification enables normalization of 2D object features to be robust to view angle changes. With



(a) Size before rectification (b) Size after rectification

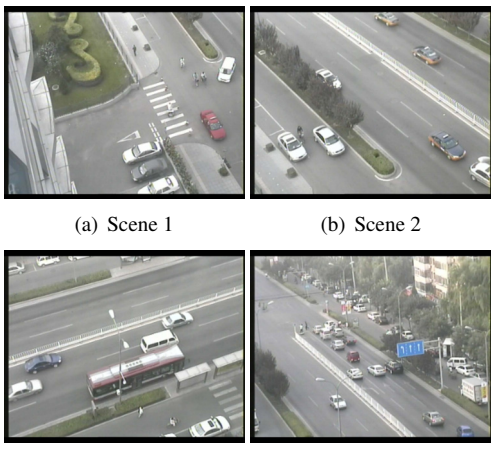
(c) Speed before rectification (d) Speed after rectification

(e) Compactness before rectification (f) Compactness after rectification

Figure 10. Effect of feature rectification

abundant moving objects extracted by motion detection and labeled manually, we analyze the class-conditional densities of the three object features before and after normalization as shown in Figure 10. As we can see, the rectified features are much easier to be classified than original features. Further more, the class-conditional densities of these three features approximately satisfy Gaussian distributions, which confirms the Gaussian assumption described in Section 5 to a certain extent.

6.3 Classification Performance



(a) Scene 1 (b) Scene 2

(c) Scene 3 (d) Scene 4

Figure 11. Illustration of traffic scenes

We test the performance of our approach in four traffic scenes of different view angles as shown in Figure 11 to test the performance of the proposed object classification approach. The average classification accuracy of the proposed method is shown in Table 1.

Table 1. Classification confusion matrix using rectified features

Accuracy	Pedestrians	Bicycles	Vehicles
Pedestrians	99.1%	0.9%	0.0%
Bicycles	2.1%	94.7%	3.2%
Vehicles	0.0 %	1.8%	98.2 %

In contrast, if we only realize classification based on original features without rectification, the classification performance is shown in Table 2. As we can see, feature rectification can greatly boost the performance of object classification.

Table 2. Classification confusion matrix using distance from cluster

Accuracy	Pedestrians	Bicycles	Vehicles
Pedestrians	88.6%	9.6%	1.8%
Bicycles	12.7%	75.5%	10.8%
Vehicles	5.2%	10.5%	84.3%

Further more, we also compare the performance of the method described in [15], with classification accuracy shown in Table 3. Evidently, the performance of our method is still better than the method of [15]. That is because the ground plane rectification can deal with perspective distortion much better than simple scene division.

Table 3. Classification confusion matrix using Gaussian Assumption

Accuracy	Pedestrians	Bicycles	Vehicles
Pedestrians	98.2%	1.8%	0.0%
Bicycles	3.4%	90.4%	6.2%
Vehicles	0.0 %	2.7%	97.3 %

6.4 Discussion

In our approach, initialization and refinement of classifiers are carried out online. With the scene change detection modular, our approach can detect scene changes and adapt to new scenes automatically. Conventional object tracking

can improve the performance of object classification with temporal information based on decision level fusion. In our implementation, our algorithm can deal with videos with the speed of 15 frames per second, which basically achieves real-time performance.

There are still many other applications of the ground plane rectification like camera calibration. Instead of a tiring wide site survey , complete calibration can now be achieved only by measuring few metrics like a horizonal line length and the camera height, which greatly decreases the difficulty for camera calibration.

The degenerate case of our approach corresponds to the top-down camera view. However, perspective distortion is not evident in top-down views. Since the ground plane rectification is taken to deal with perspective distortion, top-down view is not a big problem of our approach. Further more, traffic scene surveillance prefers to mount cameras with an oblique to the ground plane to obtain a wider view field.

From the above, we can see that our object classification algorithm has many desirable properties. It is efficient to be real-time, effective and view independent. Further more, the algorithm is free from manual labeling and supervised learning, and can deal with environment changes very well, which has great potential to be applied in real applications.

7. Conclusions

In this paper, we have proposed an approach for view independent object classification based on automated ground plane rectification. With the ground plane rectification achieved based on appearance and motion information of extracted moving vehicles, the 2D features are rectified and organized efficiently for classification. Using a novel classification framework, the classifiers are initialized and refined online and free of manual labeling. The algorithm is effective and robust to condition changes, which has great potential to be applied in real applications.

Acknowledgement

This work is funded by research grants from the National Basic Research Program of China (2004CB318110), the National Science Foundation (60605014, 60332010, 60335010 and 2004DFA06900). The authors also thank the anonymous reviewers for their valuable comments.

References

[1] B. Bose and E. Grimson. Ground plane rectification by tracking moving objects. In *Proceedings of the Joint International Workshops on VS-PETS*, 2003.

[2] L. M. Brown. View independent vehicle/person classification. In *Proc. of the ACM 2nd international workshop on Video Surveillance and Sensor Networks*, 2004.

[3] J. Deutscher, M. Isard, and J. MacCormick. Automatic camera calibration from a single manhttan image. In *Proceedings of ECCV*, 2002.

[4] D.Liebowitz and A.Zizzerman. Metric rectification for perspective images of planes. In *Proceedings of IEEE Conference on CVPR*, 1998.

[5] E.Rivlin, M.Rudzsky, R. Goldenberg, U. Bogomolov, and S. Lepchev. A real-time system for classfication of moving objects. In *Proceedings of the 16th International Conference on Pattern Recognition*, 2002.

[6] M. Everingham, A. Zisserman, C. K. I. Williams, and ... The 2005 pascal visual object classes challenge. *Lecture Notes in Computer Science*, 3944:117–176, Jan. 2006.

[7] W. Grimson, L. Lee, R. Romano, and C. Stauffer. Using adaptive tracking to classify and monitor activities in a site. In *in IEEE Conference on Computer Vision and Pattern Recognition*, 1998.

[8] F. Han, Y. Shan, R. Cekander, H. S. Sawhney, and R. Kumar. A two-stage approach to people and vehicle detection with hog-based svm. In *Performance Metrics for Intelligent Systems 2006 Workshop*, 2006.

[9] D. Liebowitz, A. Criminisi, and A. Zisserman. Creating architecural models from images. In *Proceedings of EuroGraphics*, 1999.

[10] Z. Liu, K. Huang, and T. Tan. Cast shadow removal with gmm for surface reflectance component. In *Proceedings of ICPR*, 2006.

[11] F. Lv, T. Zhao, and R. Nevatia. Using vanishing points for camera calibration. *IEEE Transactions on PAMI*, 28(9), 2006.

[12] R.I.Hartley and Zizzerman. Multiple view geometry in computer vision. *Cambridge University Press.*, 2000.

[13] C. Stauffer, K. Tieu, and L. Lee. Robust automated planar normalization of tracking data. In *Proceedings of VS-PETS 2003*, 2003.

[14] P. A. Viola, M. J. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. In *Proceedings of 9th IEEE International Conference of Computer Vision*, 2003.

[15] Z. Zhang, Y. Cai, K. Huang, and T. Tan. Real-time moving object classification with automatic scene division. In *IEEE Conference of Image Processing 2007*, 2007.

[16] Z. Zhang, M. Li, K. Huang, and T. Tan. Robust automated ground plane rectification based on moving vehicles for traffic scene visual surveillance. In *Proceedings of IEEE Conference on Image Processing*, 2008.

[17] Q. Zhou and J.K.Aggarwal. Tracking and classifying moving objects from video. In *Proc. of 2nd IEEE International Workshop on PETS*, 2001.