International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.6, No.5 (2013), pp.129-142
http://dx.doi.org/10.14257/ijsip.2013.6.5.12

Robust Vehicle Registration Method Based on 3D Model for Traffic Surveillance Application

Yuan Zheng* and Silong Peng

Institute of Automation, Chinese Academy of Sciences, Beijing, China
yuan.zheng@ia.ac.cn, silong.peng@ia.ac.cn

Abstract

3D-2D vehicle registration provides a new way for vehicle recognition, localization and tracking in traffic surveillance systems. In this paper we present two novel fitness functions to measure 3D-2D vehicle matching, where 3D wire-frame model is used. Unlike the existing vehicle registration methods, we group the model's wireframes into the important and unimportant ones in view of the disaccord between real vehicle and the wire-frame model. The important wireframes generally well fit the corresponding image edges, whereas the image edges corresponding to the unimportant wireframes may not exist due to the streamlined design of real vehicle. For more accurate matching, the fitting of the important wireframes is underlined in both two fitness functions. In the first fitness function, the larger weight coefficient is assigned to the fitting of the important wireframes; in the second fitness function, two different functions are used for the fitting of the model's wireframes instead of two different weight coefficients. Experiments on real traffic videos verify the correctness and robustness of the proposed fitness functions.

Keywords: *3D-2D matching, wire-frame model, local image gradient, traffic surveillance*

1. Introduction

3D-2D object registration has become an active research field in computer vision and image processing, which can be used for object recognition [1] and tracking [2]. For a target object, the 3D-2D registration is to match its 3D information with its 2D image information, where the 3D information includes 3D model, range data and LiDAR data. When the object's 3D model is known as a prior knowledge, the 3D-2D registration generally is converted into 2D-2D matching problem, which means converting 3D information to 2D information and then performing the matching in 2D image domain. In this kind of methods, the features of 3D model, such as points [3-5] and lines [1][2], are projected onto the image plane and then they are matched with the corresponding image features. When the object's 3D data such as range data and LiDAR data is known as a prior knowledge, the 3D-2D registration generally is converted into 3D-3D matching problem, which means converting 2D information to 3D information and performing the matching in 3D space domain. Zhao *et al.*, [6] computed 3D point clouds from continuous video and employed ICP (Iterative Closest Point) algorithm to align them with the point clouds directly obtained from LiDAR scanner. Instead of the dense point clouds, Smith *et al.*, [7] backprojected the SIFT keypoints on the image to form the new keypoints in 3D space and then matched these new keypoints with range data.

For traffic surveillance systems, moving vehicles are the target objects. In recent years, 3D-2D vehicle registration has attracted more and more attentions, which provides a new way for vehicle recognition [8-11], localization [12-14] and tracking [15-18]. With the rapid development of 3D modeling technology, 3D vehicle model can be easily obtained. The

ISSN: 2005-4254 IJSIP

model-based vehicle registration is also converted to 2D-2D matching problem. That is to say, the matching in 2D image domain is performed after projecting 3D vehicle model onto the image plane.

1.1. Related Work

For model-based vehicle registration methods, the key is to construct a fitness function that evaluates the goodness-of-fit between model projection and image data. According to the image information used for constructing the fitness function, the model-based vehicle registration methods are roughly divided into four categories: edge-based, contour-based, segmentation-based and intensity-based.

The edge-based registration methods are to extract image edges and use them to construct the fitness function. In [15, 18], the authors performed edge extraction and correspondence, and then took the distance between model projection and the corresponding image edges as the fitness function. Lou *et al.*, [16] and Wiedemann *et al.*, [2] sampled points on the image edges and regarded the distance between image edge points and model projection as the fitness function. Instead of the Euclidean distance, an improved weighted square Hausdorff distance [19] is used to measure the matching degree between image edges and model projection, where finding edge correspondences is not required. However, accurate edge extraction and distance calculation are time-consuming and sensitive to clutter and occlusion.

In the contour-based registration methods, the fitness function is defined as the similarity measure between the projected model contour and the image silhouette obtained by foreground segmentation. The overlap area between the two contours is regarded as the similarity measure in [8, 20, 10]. For more accurate contour matching, Buch *et al.*, [8] utilized the shadow removal filter. Liebelt and Schertler [21] proposed a new similarity measure that combines the contour matching and the appearance-based mutual information measure in order to increase alignment precision. In [12], the normalized cross correlation is used to evaluate the similarity between the two contours and the contour matching is accelerated by a hierarchical clustering scheme. Nevertheless, this kind of registration methods is strongly dependent on foreground segmentation technique and is sensitive to occlusion.

The segmentation-based registration methods are to divide the image into foreground and background, and then explore the relationship between the image pixels in the neighborhood of model projection and image foreground/background. The Bayesian classification error is used in [22] to describe the relationship of image pixels around model projection with image foreground and background. Similarly, the pixel-wise posterior membership probabilities [23] is used to determine that the image pixels around model projection belong to the foreground or background. Taking the image shadow into account, Johansson *et al.*, [24] modeled the vehicle as a 3D box with box shadow, and then compared its projection with image foreground, background and shadow. However, the segmentation-based methods heavily depend on image foreground-background segmentation technique and are not robust to occlusion.

As for the intensity-based registration methods, the fitness function is to exploit image intensity or gradient information, avoiding the extraction of image feature and foreground. The intensity values of image pixels around model projection are utilized in [13]. Kollnig and Nagel [14] generated a synthetic gradient by convolving model projection with a Gaussian noise and then compared this gradient with image gradient. Brisdon firstly proposed the iconic matching method in [25], which employs the discrete derivatives of image grey values in the direction normal to the model projection. Then, Pece and Worrall [26] improved this method by adding the normalization for model projection and converting the evaluation function into a likelihood framework. The latest progress was presented in [11], where local

image gradient information around model projection is utilized. Nevertheless, this kind of methods is sensitive to image noise and sometimes is not stable.

1.2. Motivation

Analyzing the existing model-based vehicle registration methods in Section 1.1, we observe that they are all sensitive to image noise, clutter and occlusion. Unfortunately, the clutter and occlusion inevitably appear in traffic scene surveillance. Thus, it may be time-consuming and error-prone to accurately extract image edges, contour and foreground. In contrast, the computation of image intensity or gradient is relatively simple and fast. Moreover, image preprocessing using the smooth filter will reduce the influence of image noise. Inspired by the intensity-based registration methods, our motivation is to propose the novel vehicle registration method that is more accurate and robust to clutter and occlusion.

In the existing vehicle registration methods, the wire-frame model is widely used due to its simplicity. Currently, more real vehicles adopt the streamlined design and thus the disaccord between real vehicle and the wire-frame model is more obvious. We discover that the image edges corresponding to some model's wireframes may not exist due to the streamlined design, particularly in the front part or rear part of the vehicle (see Figure 1). This fact implies that the goodness-of-fit score of these wireframes has low reliability. In view of this, we group the model's wireframes into important and unimportant ones, where the important wireframes generally well fit the corresponding image edges and the unimportant wireframes are just the opposite. It is a natural idea to emphasize the fitting of the important wireframes in order to obtain more accurate vehicle matching and overcome the effect of the outlier (*e.g.*, clutter and occlusion). Based on this idea, we propose two new fitness functions to evaluate 3D-2D vehicle matching. By contrast, the existing model-based methods ignore the disaccord between real vehicle and 3D wire-frame model and equally treat the fitting of the model's wireframes.

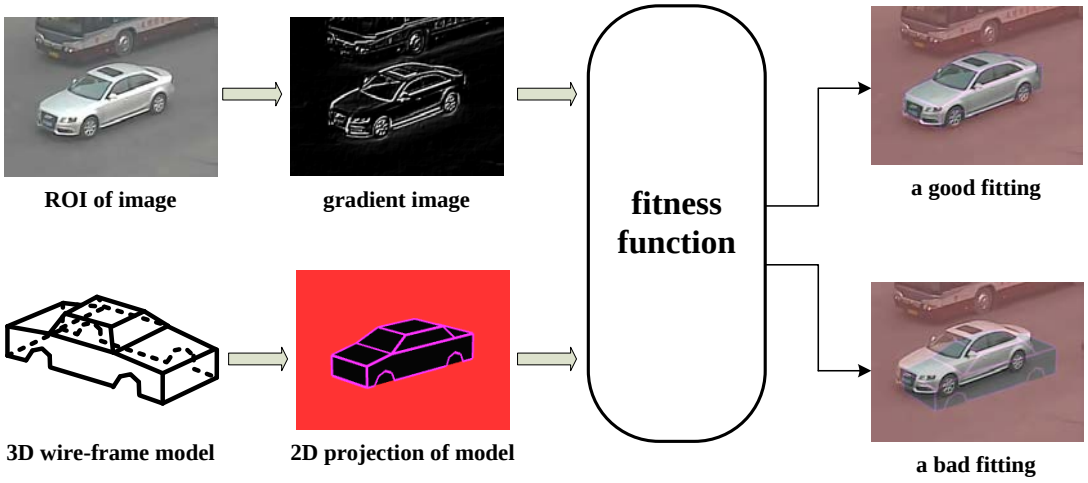


Figure 1. Illustration of 3D-2D Vehicle Registration Problem

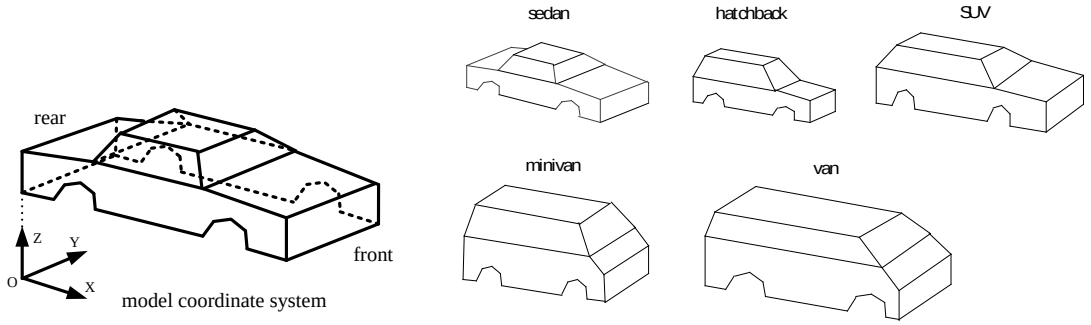
The image gradient information is exploited in both two fitness functions. To underline the fitting of the important wireframes, the first fitness function is to assign larger weight coefficient to the fitting of the important wireframes and smaller weight coefficient to the fitting of the unimportant wireframes; in the second fitness function, two different functions are used for the fitting of the model's wireframes instead of two different weight coefficients.

To improve the accuracy of the fitness function, we incorporate the normalization for the projection of the wire-frame model into these two fitness functions.

The rest of the paper is organized as follows: Section 2 presents the preliminary knowledge about vehicle wire-frame model and its projection. Section 3 proposes two novel fitness functions to evaluate 3D-2D vehicle matching. Experimental results on real traffic videos are given in Section 4. Some concluding remarks are presented in Section 5.

2. Vehicle Wire-frame Model and its Projection

The wire-frame model is widely used in the model-based vehicle registration methods. The vehicle wire-frame model consists of several 3D line segments which describe the vehicle outline and the borders with high boundary contrast (*e.g.*, the edges of vehicle window). In real traffic scenes, the vehicles of different types may appear. Given the size of each vehicle type, the wire-frame model is accordingly built. We set up a database of the vehicle wire-frame models which includes sedan, hatchback, van, minivan and SUV (see Figure 2(b)). The advantage of using 3D model consists in the robustness against the variations in viewpoint, illumination and color.



(a) The Wire-frame Model of a Sedan (b) Several Vehicle Wire-frame Models

Figure 2. Vehicle Wire-frame Model

Because the task of the fitness function is to compare model projection with image data, the 2D projection of 3D wire-frame model is certainly required. According to camera imaging principle, the mapping from model coordinate system (MCS) to image coordinate system (ICS) is expressed as

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = M_{cam} \cdot M_{pose} \cdot \begin{pmatrix} X_{\mathrm{mod}} \\ Y_{\mathrm{mod}} \\ Z_{\mathrm{mod}} \\ 1 \end{pmatrix}, \tag{1}$$

where $(X_{\mathrm{mod}}, Y_{\mathrm{mod}}, Z_{\mathrm{mod}})$ denotes a vertex of 3D wire-frame model in MCS, (u, v) denotes its projected point in ICS, λ is a scale factor, M_{cam} is the 3×4 camera projection matrix and M_{pose} is the 4×4 pose matrix that describes the rigid transformation of 3D wire-frame model from MCS to world coordinate system (WCS). Taking a sedan as an example, Figure 2(a) shows its wire-frame model in MCS. For the vehicles of different types, the vertex coordinates of their wire-frame model in MCS differ. That is to say, the vector of $(X_{\mathrm{mod}}, Y_{\mathrm{mod}}, Z_{\mathrm{mod}})$ is related to the vehicle type.

From Eq. (1), we observe that a good fitting of model projection to image data can be used for: (1) vehicle type recognition, when the camera projection matrix and the vehicle spatial pose are known; (2) vehicle 3D localization, when the camera projection matrix and the vehicle type are known; (3) camera calibration, when the vehicle type and spatial pose are known.

Notice that not every projected line of 3D wire-frame model is visible on the image, since it may be occluded by the model's body from camera viewpoint. For each projected line, we need to determine its visible part and only the visible part is used for constructing the fitness function.

3. Fitness Function

The fitness function is to evaluate the goodness-of-fit between the 2D projection of 3D vehicle model and the image data of the target vehicle. Intuitively, when the projected lines of the wire-frame model coincide with the corresponding image edges, the goodness-of-fit score reaches the maximum value. At this time, the image pixels on the projected lines have the maximum gradient in the direction normal to the projected lines. Hence, the image gradient information can be used to construct the fitness function. In this section, we firstly give the computation of goodness-of-fit score of a single projected line. Then, two gradient-based fitness functions are proposed, where the model's wireframes are grouped and their fittings are differently treated by using two different weight coefficients or functions.

For a visible projected line, we introduce a rectangular neighborhood (see Figure 3) and compute the image gradient, normal to the direction of the projected line, at pixel points within the rectangular neighborhood. Let l_i denote the i-th visible projected line and S_{rect} denote the rectangular neighborhood of l_i. For the j-th image pixel s_j within S_{rect}, the image gradient normal to the direction of l_i is written as

$$G_{\perp l_i}\left(s_j\right) = G_{s_j} \cdot \sin\left(\beta_{s_j} - \alpha_{l_i}\right), \tag{2}$$

where G_{s_j} is the gradient magnitude of s_j, β_{s_j} is the gradient orientation of s_j and α_{l_i} is the orientation of l_i. The evaluation of goodness-of-fit within the rectangular neighborhood, $e\left(l_i\right)$, is given by

$$e\left(l_i\right) = \sum_{s_j \in S_{rect}} w_{s_j} \cdot G_{\perp l_i}^2, \tag{3}$$

where w_{s_j} is the weight of s_j. w_{s_j} obeys Gaussian distribution as follows

$$w_{s_j} = \frac{1}{\omega\sqrt{2\pi}} \exp\left(-\frac{d^2}{2\omega^2}\right), \tag{4}$$

where d is the distance from s_j to l_i in pixels and ω is the half width of S_{rect}. Obviously, the image pixels that are closer to the projected line make greater contribution for constructing the fitness function. From Eq. (3), it can be seen that $e\left(l_i\right)$ is the weighted sum of the gradient information. Its weighted average can be derived as

$$\frac{\sum_{s_j \in S_{rect}} w_{s_j} \cdot G_{\perp l_i}^2}{\sum_{s_j \in S_{rect}} w_{s_j}} = \frac{\sum_{s_j \in S_{rect}} w_{s_j} \cdot G_{\perp l_i}^2}{\sum_{m=1}^{L_i} \sum_{s_j \in Q_m} w_{s_j}} = \frac{\sum_{s_j \in S_{rect}} w_{s_j} \cdot G_{\perp l_i}^2}{\sum_{m=1}^{L_i} c} \propto \frac{\sum_{s_j \in S_{rect}} w_{s_j} \cdot G_{\perp l_i}^2}{L_i}, \quad (5)$$

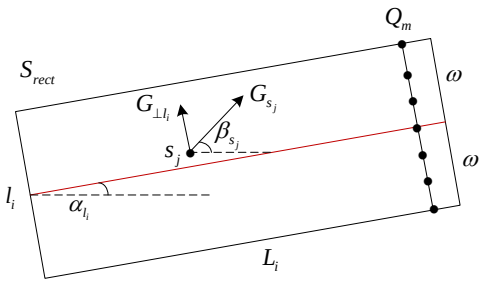


Figure 3. Rectangular Neighborhood

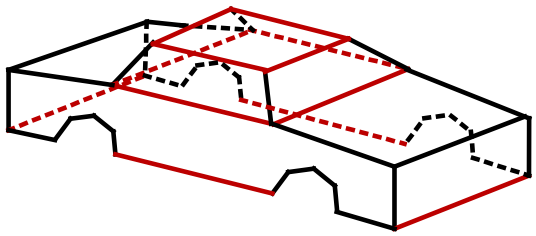


Figure 4. Model's Wireframe Grouping

where Q_m is the m-th normal of l_i within S_{rect} and L_i denotes the length of l_i in pixels. For the same ω, $\sum_{s_j \in Q_m} w_{s_j}$ is a constant, where $m = 1, \cdots, L_i$. It is found that the weighted average of $e(l_i)$ is equivalent to the normalization for the length of the projected line l_i.

Based on the above derivation, we define the normalized measure of goodness-of-fit for the visible projected line l_i as

$$M(l_i) = \sqrt{\frac{\sum_{s_j \in S_{rect}} w_{s_j} \cdot G_{\perp l_i}^2}{L_i}} = \sqrt{\sum_{s_j \in S_{rect}} \frac{w_{s_j}}{L_i} \cdot G_{\perp l_i}^2}, \quad (6)$$

The better l_i fits the corresponding image edge, the greater the value of $M(l_i)$ is. Unlike [11, 26, 25], the goodness-of-fit formula of the projected line, $M(l_i)$, is the weighted 2-norm of image gradient information rather than the weighted 1-norm of image gradient information.

Considering the disaccord between the wire-frame model and real vehicle, we group the model's wireframes into important and unimportant wireframes. The important wireframes are defined as three sets of wireframes which represent the top, the middle and the bottom of the vehicle model shown in Figure 4 (the red lines); the rest of the wireframes is viewed as the unimportant wireframes. Notice that the important wireframes describe the vehicle borders having high boundary contrast and their corresponding image edges are available. For the unimportant wireframes, their corresponding image edges may not exist, since the streamlined design of real vehicle is to replace the wireframe with smooth surface, especially in the front part or rear part of vehicle. Accordingly, the important wireframes generally can well fit the corresponding image edges; the unimportant wireframes are subject to the outliers, which could lead to a wrong matching. That is to say, the fitting of the unimportant wireframes is less reliable than the fitting of the important wireframes. To improve the accuracy of the fitness function, a natural idea is to differently treat the fitting of the model's wireframes and underline the fitting of the important wireframes. Based on this idea, we propose two new fitness functions.

In [11, 26, 25], the same weight coefficient is assigned to the fitting of the model's wireframe. In view of the difference in the reliability of the fitting, our first fitness function is

to assign larger weight coefficient to the fitting of the important wireframes, which is expressed as

$$E = \frac{1}{N}\sum_{i=1}^{N} w_{l_i} \cdot M(l_i), \tag{7}$$

$$w_{l_i} = \begin{cases} c, & \text{when } l_i \text{ is an important wireframe} \\ 1, & \text{when } l_i \text{ is an unimportant wireframe} \end{cases}, \tag{8}$$

where N is the number of all visible projected lines, w_{l_i} is the weight coefficient of the fitting of l_i and c is a constant greater than 1. The better the goodness-of-fit between the projected lines and the corresponding image edges, the greater the value of E.

In the first fitness function, the parameter c needs to be adjusted. To avoid adjusting parameter, our second fitness function is to use two different functions for the fitting of the model's wireframes instead of two different weight coefficients. The second fitness function is given by

$$E = \frac{1}{N}\sum_{i=1}^{N} e(l_i), \tag{9}$$

$$e(l_i) = \begin{cases} \frac{1}{2}M^2(l_i), & \text{when } l_i \text{ is an important wireframe} \\ |M(l_i)|, & \text{when } l_i \text{ is an unimportant wireframe} \end{cases}, \tag{10}$$

where N is the number of all visible projected lines and $e(l_i)$ denotes an evaluation of goodness-of-fit for l_i. The better the goodness-of-fit between the projected lines and the corresponding image edges, the greater the value of E. From Eq. (10), it can be seen that two functions $\rho_1(t) = \frac{t^2}{2}$ and $\rho_2(t) = |t|$ are used for the fitting of the important wireframes and the unimportant wireframes, respectively. Notice that the function $\rho_1(t) = \frac{t^2}{2}$ is above the function $\rho_2(t) = |t|$ when $t > 2$. This means that the fitting of the important wireframes is emphasized.

From Eq. (10), we observe that the weight for the fitting of the important wireframes is $\frac{1}{2}M(l_i)$ in the second fitness function. For an important wireframe l_i, the better it fits the corresponding image edge, the greater the value of $M(l_i)$ is, and the larger weight is assigned to its fitting in the second fitness function. In contrast, the first fitness function is to assign the same weight, namely parameter c, for the fitting of the important wireframes. This is the difference between these two fitness functions.

To sum up, (1) these two fitness functions perform the normalization for both length and number of the visible projected lines; (2) in view of the disaccord between real vehicle and the wire-frame model, these two fitness functions group the model's wireframes and emphasize the fitting of the important wireframes, which would improve the accuracy and robustness of the fitness function.

4. Experiments

In this section, experiments on real traffic surveillance videos are performed to verify the performance of the proposed fitness functions. We develop a software platform using OpenCV library and OpenSceneGraph(OSG) library. By means of this platform, we can create 3D wire-frame model of vehicles, simulate the roadside camera and obtain the projection of the wire-frame model. Taking real traffic image as the background image, this platform helps us visually evaluate the goodness-of-fit between model projection and image data. For experimental analysis, we choose four typical traffic videos under different traffic scenes, camera viewpoints, weather conditions and image qualities shown in Figure 5, where the red arrow denotes the positive direction of the road lane to be monitored.

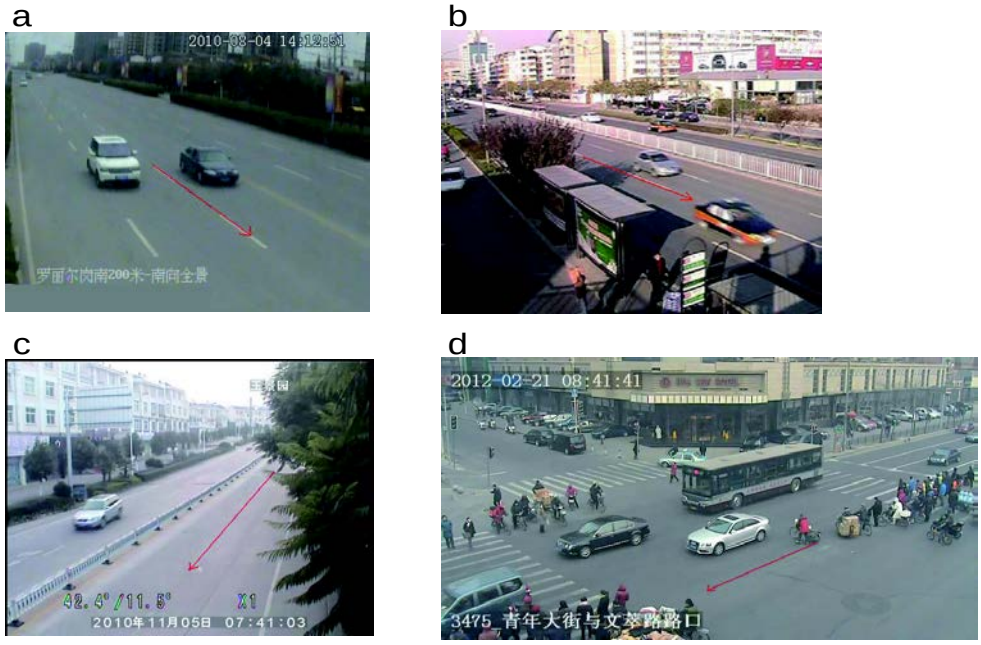


Figure 5. Test Videos under Various Conditions. (a) Traffic Scene 1: Left Viewpoint, Cloudy and Blurred Image. (b) Traffic Scene 2: Left Viewpoint, Sunny and Clear Image. (c) Traffic Scene 3: Right Viewpoint, Sunny and Blurred Image. (d) Traffic Scene 4: Right Viewpoint, Cloudy and Clear Image

The proposed fitness functions exploit image gradient information. In view of this, the Bilateral filter is used to remove image noise and the Sobel operator is used to calculate both gradient magnitude and orientation. The parameter ω in the fitness function can be roughly calculated according to a principle of similar triangles

$$\frac{\omega}{f} = \frac{\Delta d}{d_{cam-obj}}, \tag{11}$$

where f is the camera focal length in pixels, $d_{cam-obj}$ is the rough distance from the camera to the target vehicle in mm and Δd is the expanded distance from the vehicle border (usually we choose Δd = 100 mm).

Here the camera that captures traffic surveillance images is calibrated beforehand. By inputting known camera parameters, our software platform is able to simulate this roadside camera. Since the vehicle moves on the ground plane, its 3D pose is reduced to three degrees

of freedom, namely the position on the ground plane and the orientation formed by rotating about the axis normal to the ground plane. In order to verify the correctness and robustness of the fitness function, we compute the fitness score (the value of the fitness function) within the range of both position (the range of position can be visually determined by using our software platform) and orientation of the target vehicle. By sorting these fitness scores, the pose corresponding to the maximum fitness score is obtained. Under this pose, if the model projection well fits the image data of the target vehicle, it means that the fitness function is reasonable and correct; otherwise, it means that the fitness function is unable to properly evaluate the vehicle matching degree. Here we compare these two proposed fitness functions with iconic method [25] and Zhang's method [11]. Notice that: (1) both iconic method and Zhang's method do not differently treat the fitting of the model's wireframes; (2) iconic method performs the normalization for the number of sample points, whereas Zhang's method does not perform the normalization for model projection.

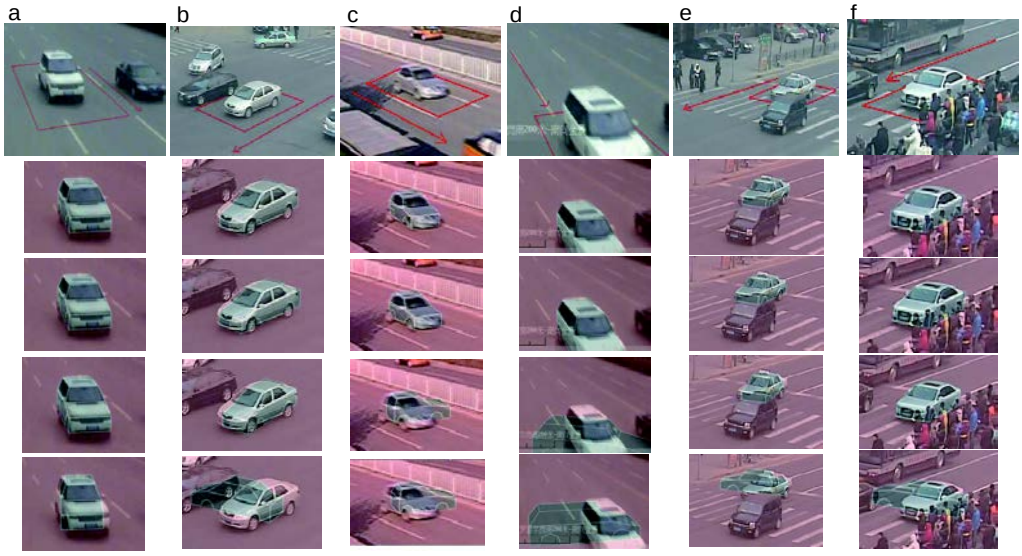


Figure 6. Several Matching Results of Different Methods. The First Row is the ROI of Test Images. The Second Row to the Fifth Row Give the Model Projection under the Pose Corres- Ponding to the Maximum Fitness Score using our First Fitness Function, our Second Fitness Function, Iconic Method and Zhang's Method, Respectively. Notice that ω=5, 4, 3, 10, 3, 3 in (a)~(f), Respectively

Three cases are discussed: (1) general case, namely without obvious clutter and occlusion; (2) when the clutter exists; (3) when the occlusion exists. In test videos, we choose the vehicles of different colors and types as target vehicles, including sedan, hatchback, van, minivan and SUV. Given the type of the target vehicle, the corresponding wire-frame model is selected from the database of vehicle model. For each case, we conduct experiments on 80 matching instances under different target vehicles and image frames. Under the pose that corresponds to the maximum fitness score, we project 3D vehicle model and observe the matching degree using our software platform.

If the model projection visually fits the image data of the target vehicle very well, a correct vehicle matching is obtained. Table 1 lists the number of the correct matching using different methods under three cases. From this table, it can be seen that: (1) in the general case, our two fitness functions slightly outperform the other two methods; (2) when the clutter or occlusion

exist, our two fitness functions obviously outperform the other two methods. This fact reveals that the proposed fitness functions are more accurate and robust to clutter and occlusion, which means that the emphasis on the fitting of the important wireframes and the normalization for model projection are helpful to improve the accuracy and robustness of the fitness function.

Table 1. Corret Matching Number of Different Methods

Method	General case		Case of clutter		Case of occlusion	
	ACM[1]	RCM[2]	ACM	RCM	ACM	RCM
Iconic	66	82.5%	54	67.5%	56	70%
Zhang's	68	85%	44	55%	52	65%
Our first[3]	74	92.5%	72	90%	70	87.5%
Our second	76	95%	72	90%	72	90%

[1]Here ACM stands for the amount of the correct matching.
[2]Here RCM stands for the rate of the correct matching.
[3]Here the parameter c=5 is used.

From these matching instances, we select several typical instances under general case, the case of clutter and the case of occlusion shown in Figure 6, where the position range of the target vehicle is marked by red frame. In Figure 6(a) and (b), there is no obvious clutter and occlusion. In Figure 6(a), the target vehicle is a white SUV. When using our two fitness functions and iconic method, the projection of the vehicle model well fits the image data of the target vehicle. When using Zhang's method, the orientation is inaccurate, although the position is accurate. In Figure 6(b), the target vehicle is a silver sedan and our second fitness function obtains the best matching result.

In Figure 6(c) and (d), the target vehicles are a gray hatchback and a white SUV respectively. Notice that the white fences on the road in Figure 6 (c) are viewed as clutter; in Figure 6 (d), the background words in the lower-left of the image are viewed as clutter and the target vehicle is partially occluded by image boundary. As can be seen, only our two fitness functions obtain the correct vehicle matching. In Figure 6 (e), the target vehicle is a silver taxi and it is partially occluded by a moving vehicle. Although the position obtained using iconic method is accurate, the orientation deviates from its ground truth. When using Zhang's method, the model projection obviously does not fit the image data. In Figure 6 (f), the target vehicle is a gray sedan and it is partially occluded by the static pedestrians. Our two fitness functions and iconic method obtain the satisfactory matching results, whereas Zhang's method does not.

Next, we discuss how to select the parameter c in the first fitness function. When the value of the parameter c is 5, 10, 50, 100, 300 and 500, we conduct experiments on some matching instances including general case, the case of clutter and the case of occlusion. If a correct vehicle matching is available when $c=5$, the satisfactory matching results still are obtained when c =10,50,100,300 or 500; if a correct vehicle matching is unavailable when $c=5$, the satisfactory matching results also are not obtained when c =10,50,100,300 or 500. This experimental results show that the value of the parameter c has little impact on matching result when $c \geq 5$.

Finally, we discuss the computational cost of the proposed fitness functions. Since the fitness function is frequently used in vehicle recognition or localization, its computational

cost is closely related to the performance of recognition and localization. The difference between our two fitness functions is that the weight for the fitting of the important wireframes differs. As a result, the computational time of these two fitness functions is almost the same and we take the second fitness function as an example. When calculating the fitness scores within the pose range of the target vehicle, we record the computational time consumed by the second fitness function. Under different poses, the computational time differs because the number and length of the visible projected lines are different. Table 2 lists the maximum, minimum and average values of the computational time using our second fitness function. More visible projection, longer projected line and more image pixels having gradient values within the rectangular neighborhood may lead to longer computational time. From Table 2, we observe that the average computation time is far less than one second, which demonstrates that our fitness function is low time-consuming.

Table 2. Computational Time of our Second Fitness Function (in ms)

	Figure 8(a)	Figure 8(c)	Figure 8(f)
Maximum	109	94	156
Minimum	46	47	109
Average	64	63	126

To sum up, we conduct extensive experiments on real traffic videos and the experimental results show that our two fitness functions are: (1) more accurate under different traffic scenes, camera viewpoints, weather conditions and image qualities; (2) more robust to clutter and partial occlusion, where the partial occlusion includes the occlusion by image boundary, moving objects and static objects; (3) low time-consuming.

5. Conclusion

In this paper, we have proposed two fitness functions based on local image gradient to evaluate 3D-2D vehicle registration. Considering the disaccord between real vehicle and 3D wire-frame model of vehicle, the model's wireframes are grouped into the important and unimportant ones. To underline the fitting of the important wireframes, two different weight coefficients and functions are used for the fitting of the model's wireframes in the first and second fitness functions, respectively. In addition, we have incorporated the normalization for the length and number of the projected lines into the fitness function. Experimental results on real traffic surveillance videos reveal that the two proposed fitness functions are more correct and robust to clutter and occlusion than the state-of-the-art registration methods. Hence, these two fitness functions can be used in more traffic scenes even the challenging scenes. In future work, the proposed vehicle registration method can be used for vehicle type recognition, 3D localization and tracking.

Acknowledgements

The authors want to thank the Beijing ViSystem Company for providing extensive real traffic surveillance videos and thank the reviewers for their thoughtful comments. This work is supported by the State Key Program of National Natural Science of China (Grant No.61032007) and the State Key Program of National Natural Science of China (Grant No.U1201251).

References

[1] P. David and D. DeMenthon, "Object recognition in high clutter images using line features", 10th IEEE International Conference on Computer Vision (ICCV), vol. 2, **(2005)**, pp. 1581-1588.
[2] C. Wiedemann, M. Ulrich and C. Steger, "Recognition and tracking of 3d objects", Pattern Recognition, **(2008)**, pp. 132-141.
[3] A. D. Bue, "Adaptive metric registration of 3d models to non-rigid image trajectories", Computer Vision–ECCV 2010, **(2010)**, pp. 87-100.
[4] S. Hinterstoisser, S. Benhimane and N. Navab, "N3m: Natural 3d markers for real-time object detection and pose estimation", 11th IEEE International Conference on Computer Vision (ICCV), **(2007)**, pp. 1-7.
[5] C. Meilhac and C. Nastar, "Robust fitting of 3d cad models to video streams", Image Analysis and Processing, **(1997)**, pp. 661-668.
[6] W. Zhao, D. Nister, and S. Hsu, "Alignment of continuous video onto 3d point clouds", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, **(2005)**, pp. 1305-1318.
[7] E. R. Smith, B. J. King, C. V. Stewart and R. J. Radke, "Registration of combined range–intensity scans: Initialization through verification", Computer Vision and Image Understanding, vol. 110, no. 2, **(2008)**, pp. 226-244.
[8] N. B. J. Orwell and S. A. Velastin, "Urban road user detection and classification using 3d wire frame models", IET Computer Vision, vol. 4, no. 2, **(2010)**, pp. 105-116.
[9] Y. Guo, C. Rao, S. Samarasekera, J. Kim, R. Kumar and H. Sawhney, "Matching vehicles under large pose transformations using approximate 3d models and piecewise mrf model", IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008), **(2008)**, pp. 1-8.
[10] S. Messelodi, C. M. Modena and M. Zanin, "A computer vision system for the detection and classification of vehicles at urban road intersections", Pattern Analysis & Applications, vol. 8, no. 1, **(2005)**, pp. 17-31.
[11] Z. Zhang, T. Tan, K. Huang and Y. Wang, "Three-dimensional deformable-model-based localization and recognition of road vehicles", IEEE Transactions on Image Processing, vol. 21, no. 1, **(2012)**, pp. 1-13.
[12] C. Reinbacher, M. R¨uther and H. Bischof, "Pose estimation of known objects by efficient silhouette matching", Proceedings of the 2010 20th International Conference on Pattern Recognition, **(2010)**, pp. 1080-1083.
[13] T. Tan and K. D. Baker, "Efficient image gradient based vehicle localization", IEEE Transactions on Image Processing, vol. 9, no. 8, **(2000)**, pp. 1343-1356.
[14] H. Kollnig and H. H. Nagel, "3d pose estimation by directly matching polyhedral models to gray value gradients", International Journal of Computer Vision, vol. 23, no. 3, **(1997)**, pp. 283-302.
[15] M. Haag and H. H. Nagel, "Combination of edge element and optical flow estimates for 3d-model-based vehicle tracking in traffic image sequences", International Journal of Computer Vision, vol. 35, no. 3, **(1999)**, pp. 295-319.
[16] J. Lou, T. Tan, W. Hu, H. Yang and S. J. Maybank, "3-d model-based vehicle tracking", IEEE Transactions on Image Processing, vol. 14, no. 10, (2005), pp. 1561-1569.
[17] A. Ottlik and H. H. Nagel, "Initialization of model-based vehicle tracking in video sequences of inner-city intersections", International Journal of Computer Vision, vol. 80, no. 2, **(2008)**, pp. 211-225.
[18] D. Roller, K. Daniilidis and H. H. Nagel, "Model-based object tracking in monocular image sequences of road traffic scenes", International Journal of Computer Vision, vol. 10, no. 3, **(1993)**, pp. 257-281.
[19] B. Yan, S. Wang, Y. Chen and X. Ding, "Deformable 3-d model based vehicle matching with weighted hausdorff and eda in traffic surveillance", 2010 International Conference on Image Analysis and Signal Processing (IASP), **(2010),** pp. 22-27.
[20] S. Gupte, O. Masoud, R. F. K. Martin and N. P. Papanikolopoulos, "Detection and classification of vehicles", IEEE Transactions on Intelligent Transportation Systems, vol. 3, no. 1, **(2002)**, pp. 37-47.
[21] J. Liebelt and K. Schertler, "Precise registration of 3d models to images by swarming particles", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), **(2007)**, pp. 1-8.
[22] Q. Liu, J. Lou, W. Hu and T. Tan, "Pose evaluation based on bayesian classification error", Proc. of 14th British Machine Vision Conference, **(2003)**, pp. 409-418.
[23] V. A. Prisacariu and I. D. Reid, "Pwp3d: Real-time segmentation and tracking of 3d objects", International journal of computer vision, vol. 98, no. 3, **(2012),** pp. 335-354.
[24] B. Johansson, J. Wiklund, P. E. Forss´en and G. Granlund, "Combining shadow detection and simulation for estimation of vehicle size and position", Pattern Recognition Letters, vol. 30, no. 8, **(2009)**, pp. 751-759.
[25] K. Brisdon, "Hypothesis verification using iconic matching", PhD thesis, University of Reading, **(1990)**.
[26] A. E. C. Pece and A. D. Worrall, "Tracking without feature detection", IEEE Int. Workshop Performance Evaluation of Tracking and Surveillance, **(2000)**, pp. 29-37.

Authors

Yuan Zheng received her Bachelor's and Master's degrees from Harbin Institute of Technology (HIT), Harbin, China, in 2006 and 2008, respectively. She is currently a Ph.D. candidate at the Institute of Automation, Chinese Academy of Sciences, Beijing. Her research interests include computer vision, image and video analysis.
E-mail: yuan.zheng@ia.ac.cn

Silong Peng received the bachelor’s degree in mathematics from Anhui University, Hefei, China, in 1993, and the Master’s and Ph.D. degrees in mathematics from Institute of Mathematics, Chinese Academy of Sciences, Beijing, China, in 1995 and 1998, respectively. He is currently a professor at the Institute of Automation, Chinese Academy of Sciences, Beijing. His research interests include wavelet, signal and digital image processing.
E-mail: silong.peng@ia.ac.cn