International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:10, No:6, 2016

A Practical and Efficient Evaluation Function for 3D Model Based Vehicle Matching

Yuan Zheng

Abstract—3D model-based vehicle matching provides a new way for vehicle recognition, localization and tracking. Its key is to construct an evaluation function, also called fitness function, to measure the degree of vehicle matching. The existing fitness functions often poorly perform when the clutter and occlusion exist in traffic scenarios. In this paper, we present a practical and efficient fitness function. Unlike the existing evaluation functions, the proposed fitness function is to study the vehicle matching problem from both local and global perspectives, which exploits the pixel gradient information as well as the silhouette information. In view of the discrepancy between 3D vehicle model and real vehicle, a weighting strategy is introduced to differently treat the fitting of the model's wireframes. Additionally, a normalization operation for the model's projection is performed to improve the accuracy of the matching. Experimental results on real traffic videos reveal that the proposed fitness function is efficient and robust to the cluttered background and partial occlusion.

Keywords—3D-2D matching, fitness function, 3D vehicle model, local image gradient, silhouette information.

I. Introduction

IN recent years, 3D model-based object matching is an active research area in computer vision, which is broadly used for object recognition [1]–[4], localization [5]–[8] and tracking [2], [4], [9]. For a target object, 3D model-based matching is also called as 3D-2D matching, since it is to match the 3D information with the 2D information. Specifically, it firstly projects the 3D model onto the image plane, and then matches the model projection with the corresponding image information, such as image points [10]–[14], edge [1], [2], [4], contour [5], [8], [15] and foreground information [6], [7], [9], [16].

In traffic surveillance systems, the moving vehicles in traffic scenarios are the interesting targets. With the development of 3D modeling technology, 3D vehicle model can be easily obtained. Over past decades, 3D model-based vehicle matching has attracted more and more attentions, which provides a new way for vehicle recognition [3], [17]–[19], localization [8], [19]–[21] and tracking [22]–[25]. Similarly, the 3D model-based vehicle matching is to perform the matching in 2D image domain after projecting 3D vehicle model onto the image plane.

A. Related Work

For 3D model-based vehicle matching, the key is to construct an evaluation function, also called fitness function, to measure the goodness-of-fit between the model projection and image data. According to the used image information, the existing fitness function can be broadly divided into three types as follows: The edge-based fitness function is to extract the edges from the interesting region of the image and use them to construct the fitness function. In [22], [23], the image edges are extracted and the fitness function is defined as the distance between the model projection and the corresponding image edges. In [2], [24], the image edges are sampled and the distance between the sampled points and the model projection is viewed as the fitness function. Instead of the Euclidean distance that is commonly used, [26] utilizes a weighted square Hausdorff distance and [27] utilizes a directional Chamfer distance. In [25], the authors proposed a generic 3D vehicle model where the hypothesized edges from the occluding contours and part boundaries are generated. For each hypothesized edge, the authors found the nearby image edges and used the perpendicular distance between the edge correspondence to measure the matching degree.

The silhouette-based fitness function is to exploit the silhouette information to construct the fitness function. Generally, the silhouette is obtained via the foreground and background segmentation. In this kind of methods, the fitness function is defined as the similarity measure between image silhouette and the projected model contour. In [17], [18], [28], the overlap area between these two contours is used as the similarity measure. For a more accurate contour matching, the shadow removal filter is employed in [18]. Liebelt and Schertler [5] combined the contour matching with the appearance-based mutual information measure to further increase alignment precision. In [8], the normalized cross correlation is used to evaluate the similarity of the two contours and the hierarchical clustering scheme helps accelerate the contour matching. [29] is to perform the contour matching by means of active shape model(ASM). Other works [9], [30], [31] are to construct the fitness function by exploring the relationship of the image pixels around the model projection with foreground and background (i.e., by determining that the image pixels around the model projection belong to foreground or background). Taking the shadow into account, Johansson et al. [31] modeled the vehicle as a 3D box with box shadow and compared its projection with image foreground, background and shadow.

The intensity-based fitness function is to exploit the image intensity or gradient information to construct the fitness function, avoiding the edge extraction or foreground segmentation. In [21], the intensity values of image pixels in the neighborhood of model projection are utilized. In

Y. Zheng is with the School of Computer Science, Harbin Institute of Technology Shenzhen Graduate School, China (e-mail: zhengyuan@hitsz.edu.cn).

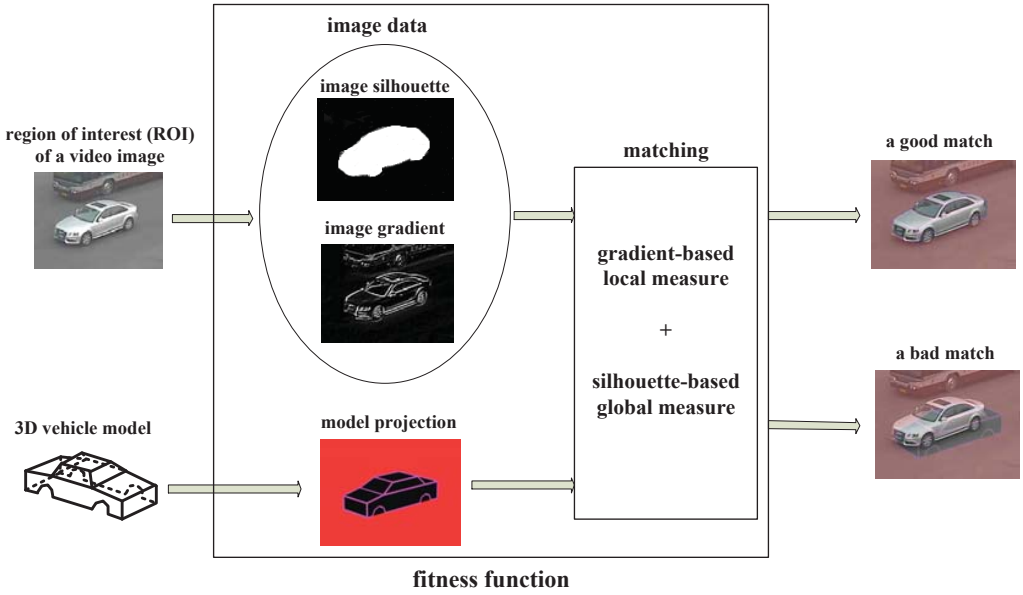


Fig. 1 Illustration of 3D model-based vehicle matching problem, where the fitness function is used to evaluate the matching between the model projection and image data: The proposed fitness function exploits the image gradient and silhouette information

practical applications, large illumination variation would lead to different intensity values of the same pixel. In view of this, illumination condition estimation is incorporated into the vehicle matching in [32], [33]. Besides pixel intensity value, the pixel gradient (i.e., the derivative of intensity value) is also exploited to construct the fitness function. In [34], the discrete derivatives of image grey values in the direction normal to the model projection are employed. Afterwards, this method is improved by adding the normalization for model projection and converting the evaluation function into a likelihood framework [35]. In [20], the authors generated the synthetic gradient by convolving the model projection with Gaussian noise, and then compared the synthetic gradient with image gradient. The latest progress was presented in [19], [36], where the local pixel gradient information around the model projection is utilized.

B. Motivation and Contributions

The existing fitness functions are generally sensitive and error-prone to outliers, such as clutter and occlusion. Nevertheless, the clutter and occlusion inevitably exist in traffic scenarios. In view of this, our motivation is to propose a novel fitness function that is more accurate and robust to clutter and occlusion.

Inspired by the intensity-based fitness functions, we also exploit the pixel gradient information since the gradient calculation is usually simple and fast. In order to further improve the accuracy of vehicle matching, except for the local gradient information, we also exploit the image silhouette information. Accordingly, we incorporate the matching between the image silhouette and the model's projected contour into the gradient-based matching. For the model-based vehicle matching, the wire-frame model of the vehicle is widely used due to its simplicity. However, there is the discrepancy between the wire-frame model and real vehicle because of the streamlined design of real vehicle. In view of this, we group the model's wireframes and present a weighting strategy to differently treat the fitting of the model's wireframes. Unlike the existing fitness functions, the proposed fitness function not only considers the local and global characteristics of vehicle matching, but also takes into account the discrepancy between 3D vehicle model and real vehicle, which would help improve its accuracy and robustness.

C. Paper Structure

The remainder of the paper is structured as follows. Section II first presents the preliminary knowledge about 3D vehicle model and its 2D projection, and then proposes a novel fitness function in view of both local and global characteristics of vehicle matching. Experiments on real traffic images are presented in Section III and some concluding remarks are presented in Section IV.

II. Methodology

A. 3D Vehicle Model and Its Projection

With the rapid development of 3D modeling technology, the 3D model of the object is available. The advantage of using 3D model consists in the robustness against the variations in viewpoint, illumination and colour. For the model-based vehicle matching, 3D wire-frame model is commonly used. It is composed of several 3D line segments that delineate the vehicle outline and some borders with high boundary contrast such as the edges of vehicle window (see Fig. 1). In real traffic scenarios, the vehicles of various types may appear. Therefore, we set up a database of the vehicle wire-frame model that includes sedan, hatchback, van, minivan and SUV.

3D model-based vehicle matching is to match the model projection and image data, and thus the 2D projection of 3D wire-frame model is certainly required. According to the camera imaging principle, the mapping from the model coordinate system(MCS) to the image coordinate system (ICS) can be expressed as

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = M_{camera} \cdot M_{pose} \cdot \begin{pmatrix} X_{model} \\ Y_{model} \\ Z_{model} \\ 1 \end{pmatrix}, \quad (1)$$

where $(X_{model}, Y_{model}, Z_{model})$ denotes a vertex of 3D wire-frame model in MCS, (u, v) is the corresponding projected point in ICS, λ is a scale factor, M_{camera} is camera projection matrix and M_{pose} is pose matrix that describes the 3D pose of the vehicle model in world coordinate system(WCS). Notice that not every projected line of 3D wire-frame model is visible, since it may be occluded by the model's body from the camera's viewpoint. For the construction of the fitness function, only the visible projected lines or the visible part of the projected line are used.

B. Fitness Function

The goal of the fitness function is to find out a best match between the image data and the projection of vehicle model. From the local perspective of vehicle matching, when the projected lines of 3D wire-frame model coincide with the corresponding image edges, a correct match is obtained. In other words, at this time the image pixels lying on the projected lines have the maximum gradient value in the direction normal to the projected lines. As a consequence, we exploit the pixel gradient information around the model projection to construct the fitness function. For every visible projected line, we introduce a rectangular neighborhood (see Fig. 2 (a)) and compute the image gradient, in the normal direction of the projected line, at the pixel points within the rectangular neighborhood. Let l_i denotes the i-th visible projected line and S_{rect} denotes the rectangular neighborhood of l_i. For the j-th image pixel s_j within S_{rect}, the image gradient perpendicular to the direction of l_i is written as

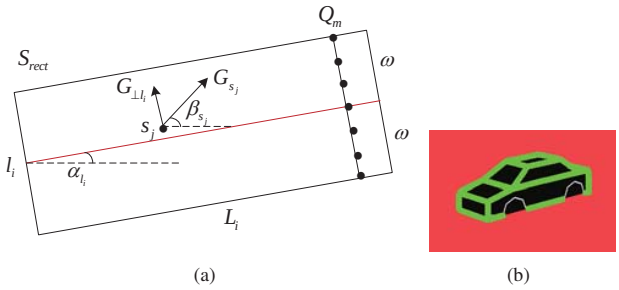
$$G_{\perp l_i}(s_j) = G_{s_j} \cdot |\sin(\beta_{s_j} - \alpha_{l_i})|, \quad (2)$$

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. Evidently, the closer the image pixel s_j is to the projected line l_i, the greater contribution it makes for constructing the fitness function. Accordingly, the weight of s_j, w_{s_j}, obeys the following Gaussian distribution

$$w_{s_j} = \frac{1}{\omega\sqrt{2\pi}} exp\left(-\frac{d^2}{2\omega^2}\right), \quad (3)$$

where d is the distance from s_j to l_i in pixels and ω is the half width of S_{rect}. For l_i, we give the normalized measure of matching degree by

$$m(l_i) = \sum_{s_j \in S_{rect}} \frac{w_{s_j}}{L_i} \cdot G_{\perp l_i}, \quad (4)$$



(a) (b)

Fig. 2 Rectangular neighborhood. (a) Rectangular neighborhood S_{rect} of the projected line l_i. Its length and width are L_i and 2ω respectively. (b) Rectangular neighborhoods of the model projection in green color

where L_i is the length of l_i in pixels. The better l_i fits the corresponding edge, the greater the value of $m(l_i)$ is.

Considering the discrepancy between 3D wire-frame model and real vehicle, we group the model's wireframes into the primary and secondary ones. The primary wireframes are defined as three sets of wireframes that delineate the top, the middle and the bottom of the vehicle (see Fig. 3 (a)); the rest of the model's wireframes is viewed as the secondary wireframes. Notice that the primary wireframes describe the vehicle borders having high boundary contrast and their corresponding image edges are available. As for the secondary wireframes, the corresponding image edges may not exist, since the streamlined design of real vehicle is to substitute the smooth surfaces for the wireframes, especially in the front part or rear part of vehicle (see Fig. 3 (b)). Consequently, the primary wireframes are able to better fit the corresponding image edges than the secondary wireframes. That is to say, the fitting of the primary wireframes has higher reliability. In order to improve the accuracy and robustness, a natural idea is to treat the model's wireframes differently and emphasize the fitting of the primary wireframes. Based on this idea, we utilize a weighting strategy for the fitting of the model's wireframes, namely

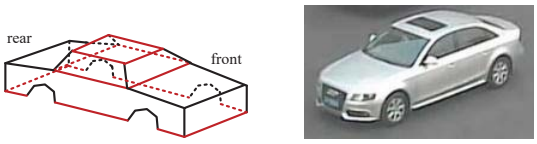
$$M(l_i) = \begin{cases} q \cdot m(l_i) & l_i \text{ is primary wireframe} \\ (1-q) \cdot m(l_i) & l_i \text{ is secondary wireframe} \end{cases}, \quad (5)$$

where q and $1-q$ are the weight coefficients for the primary wireframe and the secondary wireframe, respectively. Here we set q as a constant that is close to 1, which means that the larger weight coefficient is assigned for the primary wireframe. Taking all visible projected lines into account, we give the gradient-based measure of vehicle matching as

$$E_g = \frac{1}{n} \sum_{i=1}^{n} M(l_i), \quad (6)$$

where n is the number of the visible projected lines. The better the projected lines are aligned with the corresponding image edges, the greater the value of E_g is.

Except for the local perspective, we also consider the vehicle matching problem from the global perspective. Intuitively, if the projected contour of 3D wire-frame model exactly covers the image silhouette of target vehicle, a correct



(a) (b)

Fig. 3 Discrepancy between the vehicle model and real vehicle: (a) 3D wire-frame model and its grouping, where the primary wireframes are marked in red color, (b) Real vehicle with the streamlined design, where the smooth surfaces are substituted for the model's wireframes, especially in the front part or rear part of vehicle

match is obtained. Consequently, we utilize the overlap rate to evaluate the similarity between the projected contour and image silhouette. The silhouette-based measure for vehicle matching is given by

$$E_s = \frac{A_{model} \cap A_{image}}{A_{model} \cup A_{image}} , \tag{7}$$

where A_{model} denotes an area corresponding to model projection and A_{image} denotes an area corresponding to image silhouette. As can be seen from (7), the range of E_s is from 0 to 1. The better the contour of model projection is aligned with the corresponding image silhouette, the greater the value of E_s is.

In order to improve the accuracy and robustness to clutter and occlusion, the proposed fitness function E is to combine the gradient-based measure E_g (i.e., local measure) and the silhouette-based measure E_s (i.e., global measure), which can be expressed as

$$E = E_s \cdot E_g . \tag{8}$$

In many cases, only using the local measure is unable to obtain a satisfactory matching result, whereas incorporating the global measure can do. The advantage of incorporating the global measure into the local measure is to compensate for the weakness of the local measure and improve the accuracy of vehicle matching.

III. Experiments

In this section, the experiments on real traffic videos are conducted to verify the performance of the proposed fitness function. We develop a software platform that utilizes both OpenCV library and OpenSceneGraph(OSG) library. By means of this platform, we can create the 3D wire-frame model of vehicle, simulate the roadside camera and obtain the rendered image of 3D scene. Taking the real traffic image as the background image, this platform helps us visually evaluate the matching degree between the model projection and image data, and further helps us qualitatively evaluate the performance of the fitness function. For the experimental analysis, we select several typical traffic videos under different scenarios, camera viewpoints and weather conditions, as shown in Fig. 4. In these test videos, we choose the vehicles of different colors and types as the target vehicles, including sedan, hatchback, van, minivan and SUV. According to the type of the target vehicle, the corresponding wire-frame model is selected from the database of 3D vehicle model. In these

Fig. 4 Test video set under various conditions: (a) Traffic scene 1: right viewpoint and cloudy day, (b) Traffic scene 2: right viewpoint and sunny day, (c) Traffic scene 3: left viewpoint and cloudy day, (d) Traffic scene 4: left viewpoint and cloudy day, (e) Traffic scene 5: left viewpoint and sunny day, (f) Traffic scene 6: left viewpoint and cloudy day, (g) Traffic scene 7: right viewpoint and nighttime

experiments, we set $q = 0.8$. The parameter ω in (3) can be roughly estimated according to the following principle of similar triangles

$$\frac{\omega}{f} = \frac{\Delta d}{d_{cam-obj}} , \tag{9}$$

where f is the camera's 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 $\Delta d = 100mm$).

To verify the accuracy and robustness of the fitness function, we discuss three cases: General case (i.e., without obvious clutter and occlusion), the case of clutter and the case of occlusion. For these three cases, we conduct experiments on 220 matching instances that come from 70 vehicles of different types and colors under 7 traffic scenarios, where the matching instances for general case, the case of clutter and the case of occlusion are 100, 60 and 60 respectively. By calculating the fitness function, we first find a vehicle's pose that corresponds to the maximum matching score. Under this pose, we project 3D vehicle model and observe the matching degree using our software platform. When the model projection visually fits the image data of target vehicle very well, a correct vehicle match is obtained. Here we compare the proposed fitness function with the evaluation functions in iconic method[1] [34] and Zhang's method [19], where iconic method is a classical method and Zhang's method is the latest work. The rate of the correct match using the different evaluation functions under three cases is listed in Table I. From this table, it can be seen that: (1) in general case, our fitness function slightly outperforms the other two evaluation functions; (2) when the clutter or occlusion exist, our fitness function considerably outperforms the other two evaluation functions. That is to say, our fitness function is more accurate and robust to clutter and occlusion. From these matching instances, we select several typical instances under three cases, as shown in Fig. 5. Figs. 5 (a) and (b) give the results of vehicle matching under general case, where the target vehicles (i.e., a gray sedan and a green sedan) are not occluded at all and are not cluttered by the obvious outliers. Both our fitness function and the other two

[1]Here all pixel points within the rectangular neighborhood of the projected line are regarded as sample points and querying probability table is omitted.

Fig. 5 Examples of vehicle matching using the different evaluation functions under general case, the case of clutter and the case of occlusion: The first row gives the test images; The second, third and fourth rows give the matching results using our fitness function, the evaluation function in iconic method [34] and the evaluation function in Zhang's method [19], respectively; The parameter $\omega = 6, 4, 4, 3, 3, 10$ in (a)-(f), respectively

evaluation functions obtain the satisfactory matching result. Figs. 5 (c) and (d) give the results of vehicle matching under the case of clutter, where the target vehicles (i.e., a gray hatchback and a silver sedan) are cluttered by the white fences in the road and the roadside lawn. Figs. 5 (e) and (f) give the results of vehicle matching under the case of occlusion, where the target vehicles (i.e., a silver taxi and a white SUV) are partially occluded by a moving vehicle and the image boundary. Notice that the background words in the lower-left of the image in Fig. 5 (f) are viewed as the clutter for vehicle matching. As can be seen, when the clutter and occlusion occur, only our fitness function obtains a satisfactory matching result.

As we can see, our fitness function is capable of more accurately evaluating the vehicle matching. The reason resides in: (1) we improve the gradient-based measure by introducing a weighting strategy for the fitting of the model's wireframes and a normalization operation for the model projection; (2) we consider the global characteristics of vehicle matching by incorporating the silhouette-based measure into the gradient-based measure. Next, we test the impact of the weighting strategy, the normalization operation and the incorporation of the silhouette-based measure on the vehicle matching. Fig. 6 shows the matching results with and without both the weighting strategy and the normalization operation. From this figure, we observe that the weighting strategy and the normalization operation certainly help improve the accuracy of vehicle matching. Fig. 7 shows the matching results with and without the incorporation of the silhouette-based measure. From this figure, it can be seen that the advantage of incorporating the silhouette-based measure is to compensate for the weakness of the gradient-based measure and improve the accuracy of vehicle matching.

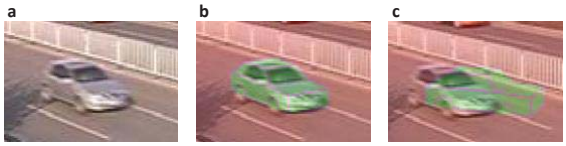


Fig. 6 Impact of both weighting strategy and normalization operation on vehicle matching: (a) Test image, (b) Matching result with the weighting strategy and normalization operation, (c) Matching result without the weighting strategy and normalization operation

Fig. 7 Impact of incorporating the silhouette-based measure on vehicle matching: (a) Test image, (b) Matching result with the incorporation of the silhouette-based measure, (c) Matching result without the incorporation of the silhouette-based measure

IV. Conclusion

In this paper, a fitness function for 3D model-based vehicle matching is presented. It takes into account the local and global characteristics of vehicle matching and makes full use of the image gradient information and silhouette information. In the gradient-based measure, considering the discrepancy between 3D vehicle model and real vehicle, the proposed fitness function treats the model's wireframes differently and emphasizes the fitting of the primary wireframes via a

TABLE I
RESULTS OF VEHICLE MATCHING USING THE DIFFERENT EVALUATION FUNCTIONS UNDER THREE CASES (TOTAL 220 MATCHING INSTANCES FROM 70 TARGET VEHICLES UNDER THE DIFFERENT TRAFFIC SCENARIOS)

Evaluation functions	Rate of the correct match		
	general case	case of clutter	case of occlusion
iconic's	92%	73.3%	80%
Zhang's	92%	75%	76.7%
ours	**95%**	**90%**	**91.7%**

weighting strategy. Except for the gradient-based measure, the proposed fitness function is to incorporate the silhouette-based measure. The advantage of this way is to compensate for the weakness of the gradient-based measure and improve the accuracy of vehicle matching. Experiments on real traffic images verify the effectiveness and practicability of the proposed fitness function.

ACKNOWLEDGMENT

The authors want to thank the Beijing ViSystem Company for providing real traffic surveillance videos and thank the reviewers for their thoughtful comments. This study was supported by the National Natural Science Foundation of China (Grant No.61502119) and the China Postdoctoral Science Foundation (Grant No.2015M571414).

REFERENCES

[1] P. David and D. DeMenthon, "Object recognition in high clutter images using line features," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 2. IEEE, 2005, pp. 1581–1588.
[2] C. Wiedemann, M. Ulrich, and C. Steger, "Recognition and tracking of 3d objects," *Pattern Recognition*, pp. 132–141, 2008.
[3] 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," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.
[4] S. Nedevschi, S. Bota, and C. Tomiuc, "Stereo-based pedestrian detection for collision-avoidance applications," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 10, no. 3, pp. 380–391, 2009.
[5] J. Liebelt and K. Schertler, "Precise registration of 3d models to images by swarming particles," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.
[6] B. Rosenhahn, T. Brox, and J. Weickert, "Three-dimensional shape knowledge for joint image segmentation and pose tracking," *International Journal of Computer Vision*, vol. 73, no. 3, pp. 243–262, 2007.
[7] S. Dambreville, R. Sandhu, A. Yezzi, and A. Tannenbaum, "Robust 3d pose estimation and efficient 2d region-based segmentation from a 3d shape prior," in *Computer Vision–ECCV 2008*. Springer, 2008, pp. 169–182.
[8] C. Reinbacher, M. Rüther, and H. Bischof, "Pose estimation of known objects by efficient silhouette matching," in *Proceedings of the 2010 20th International Conference on Pattern Recognition*. IEEE Computer Society, 2010, pp. 1080–1083.
[9] V. Prisacariu and I. Reid, "Pwp3d: Real-time segmentation and tracking of 3d objects," *International Journal of Computer Vision*, pp. 1–20, 2012.
[10] C. Meilhac and C. Nastar, "Robust fitting of 3d cad models to video streams," in *Image Analysis and Processing*. Springer, 1997, pp. 661–668.
[11] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 2. IEEE, 2005, pp. 1508–1515.
[12] S. Hinterstoisser, S. Benhimane, and N. Navab, "N3m: Natural 3d markers for real-time object detection and pose estimation," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–7.
[13] A. Del Bue, "Adaptive metric registration of 3d models to non-rigid image trajectories," *Computer Vision–ECCV 2010*, pp. 87–100, 2010.
[14] S. M. Khan, H. Cheng, D. Matthies, and H. Sawhney, "3d model based vehicle classification in aerial imagery," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 1681–1687.
[15] E. Marchand, P. Bouthemy, F. Chaumette, V. Moreau *et al.*, "Robust real-time visual tracking using a 2d-3d model-based approach," in *IEEE Int. Conf. on Computer Vision, ICCV'99*, vol. 1, 1999, pp. 262–268.
[16] C. Schmaltz, B. Rosenhahn, T. Brox, D. Cremers, J. Weickert, L. Wietzke, and G. Sommer, "Region-based pose tracking," in *Pattern Recognition and Image Analysis*. Springer, 2007, pp. 56–63.
[17] S. Messelodi, C. 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, pp. 17–31, 2005.
[18] N. Buch, J. Orwell, and S. Velastin, "Urban road user detection and classification using 3d wire frame models," *Computer Vision, IET*, vol. 4, no. 2, pp. 105–116, 2010.
[19] Z. Zhang, T. Tan, K. Huang, and Y. Wang, "Three-dimensional deformable-model-based localization and recognition of road vehicles," *Image Processing, IEEE Transactions on*, vol. 21, no. 1, pp. 1–13, 2012.
[20] H. Kollnig and H. Nagel, "3d pose estimation by directly matching polyhedral models to gray value gradients," *International Journal of Computer Vision*, vol. 23, no. 3, pp. 283–302, 1997.
[21] T. Tan and K. Baker, "Efficient image gradient based vehicle localization," *Image Processing, IEEE Transactions on*, vol. 9, no. 8, pp. 1343–1356, 2000.
[22] D. Roller, K. Daniilidis, and H. Nagel, "Model-based object tracking in monocular image sequences of road traffic scenes," *International Journal of Computer Vision*, vol. 10, no. 3, pp. 257–281, 1993.
[23] M. Haag and 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, pp. 295–319, 1999.
[24] J. Lou, T. Tan, W. Hu, H. Yang, and S. Maybank, "3-d model-based vehicle tracking," *Image Processing, IEEE Transactions on*, vol. 14, no. 10, pp. 1561–1569, 2005.
[25] M. J. Leotta and J. L. Mundy, "Vehicle surveillance with a generic, adaptive, 3d vehicle model," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 7, pp. 1457–1469, 2011.
[26] B. Yan, S. Wang, Y. Chen, and X. Ding, "Deformable 3-d model based vehicle matching with weighted hausdorff and eda in traffic surveillance," in *Image Analysis and Signal Processing (IASP), 2010 International Conference on*. IEEE, 2010, pp. 22–27.
[27] M. Hodlmoser, B. Micusik, M. Y. Liu, M. Pollefeys, and M. Kampel, "Classification and pose estimation of vehicles in videos by 3d modeling within discrete-continuous optimization," in *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*. IEEE, 2012, pp. 198–205.
[28] S. Gupte, O. Masoud, R. Martin, and N. Papanikolopoulos, "Detection and classification of vehicles," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 3, no. 1, pp. 37–47, 2002.
[29] Y. Tsin, Y. Genc, and V. Ramesh, "Explicit 3d modeling for vehicle monitoring in non-overlapping cameras," in *Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on*. IEEE, 2009, pp. 110–115.
[30] Q. Liu, J. Lou, W. Hu, and T. Tan, "Pose evaluation based on bayesian classification error," in *In Proc. of 14th British Machine Vision Conference*, 2003, pp. 409–418.
[31] B. Johansson, J. Wiklund, P. Forssén, and G. Granlund, "Combining shadow detection and simulation for estimation of vehicle size and position," *Pattern Recognition Letters*, vol. 30, no. 8, pp. 751–759, 2009.
[32] S. Jayawardena, M. Hutter, and N. Brewer, "Featureless 2d–3d pose estimation by minimising an illumination-invariant loss," in *Image and Vision Computing New Zealand (IVCNZ), 2010 25th International Conference of*. IEEE, 2010, pp. 1–8.

[33] T. Hou, S. Wang, and H. Qin, "Vehicle matching and recognition under large variations of pose and illumination," in *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*. IEEE, 2009, pp. 24–29.
[34] K. Brisdon, "Hypothesis verification using iconic matching." Ph.D. dissertation, University of Reading, 1990.
[35] A. Pece and A. Worrall, "Tracking without feature detection," in *Proc. IEEE Int. Workshop Performance Evaluation of Tracking and Surveillance*. Citeseer, 2000, pp. 29–37.
[36] M. Hodlmoser, B. Micusik, M. Pollefeys, M.-Y. Liu, and M. Kampel, "Model-based vehicle pose estimation and tracking in videos using random forests," in *3DTV-Conference, 2013 International Conference on*. IEEE, 2013, pp. 430–437.