A Low-cost Strong Shadow-based Segmentation Approach for Vehicle Tracking in Congested Traffic Scenes

Ehsan Adeli Mosabbeb[1], Maryam Sadeghi[2], Mahmood Fathy[1], Maliheh Bahekmat[3]
[1]Computer Engineering Department, Iran University of Science and Technology, Narmak, Tehran, 1684613114, Iran
[2]Computer Science Department, Simon Fraser University, Vancouver, Canada
[3]College of Engineering, Islamic Azad University of Mashhad, Mashhad, Iran
E-mail: eadeli@iust.ac.ir, msa68@sfu.ca, mahfathy@iust.ac.ir, info@bahekmat.ir

***Abstract* – In this paper, a vehicle tracking algorithm is proposed that uses a new approach to deal with occlusion. This approach uses the novel feature, recently proposed to improve the accuracy of localization and occlusion handling. It was constructed on the basis of the strong shadow under the vehicle in real-world traffic scenes. In this paper, some statistical parameters of each frame are used to detect and segment these shadows. To demonstrate robustness and accuracy of our proposed approach, impressive results of our method in real traffic images including high congestion, noise, clutter, snow, and rain containing remarkable cast shadows, bad illumination conditions and occlusions, taken from both scenes of outdoor highways and urban roads are presented.**

***Index Terms* – Background Modeling, Cast Shadow, Occlusion, Segmentation, Shadow Detection, Vehicle Tracking**

I. INTRODUCTION

Increasing congestion on freeways and problems associated with existing detectors has generated an interest in new vehicle detection and tracking technologies using computer vision techniques. Existing commercial image processing systems work well in free-flowing traffic, but the systems have difficulties with congestion, occlusion, shadows and lighting transitions. This paper addresses the problem of vehicle tracking in real-world traffic images.

A classic technique to extract the moving objects (vehicles) is background subtraction. Over the years researchers have proposed various solutions to the automated tracking problem. These approaches can be classified as follows: Blob Tracking [1]-[3], Active Contour Tracking [4], 3D-Model Based Tracking [5]-[8], Markov Random Field Tracking [9], Feature Tracking [10], [11] and Color and Pattern-Based Tracking [12]-[14]. Some of these assume an aerial view of the scene which virtually eliminates all occlusions [8]. In [7], a single vehicle is successfully tracked through a partial occlusion, but its applicability to congested traffic scenes has not been demonstrated.

In the previous work [15], a new low-cost method has been presented for occlusion handling that uses strong shadow as a key feature to vehicle detection, though shadow detection techniques have been employed for shadow removal from background and foreground. The problem of shadow detection has been increasingly addressed over the past years. Shadow detection techniques can be classified into two groups: model-based and property-based techniques. Model-based techniques are designed for specific applications, such as aerial image understanding [16] and video surveillance [17]. Luminance information is exploited in early techniques by analyzing edges [18], and texture information [19]. Luminance, chrominance and gradient density information is used in [20]. Color information is used also in [21]. A physics-based approach to distinguish material changes from shadow boundaries in chromatic still images is presented in [22]. Cavallaro et. al in [23] proposed Shadow-aware object-based video processing. A classification of color edges by means of photometric invariant features into shadow-geometry edges, highlight edges, and material changes is proposed in [24]. Using strong shadow information as a feature for vehicle detection was initially discussed in [15]. By investigating image intensity, it was found that the area under a vehicle is distinctly darker than any other areas on an asphalt paved road.

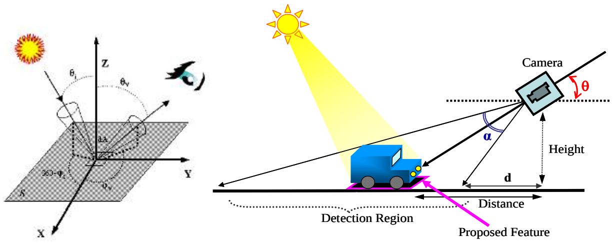
II. SHADOW ANALYSIS

A cast shadow is the area projected by the object in the direction of direct light. Shadows are characterized by two types of properties: photometric and geometric. Geometric properties depend on the type of obstruction and position of light source. Photometric properties determine relation of pixel intensity of background under illumination and under shadow. We can model geometric properties of shadow and illumination with BDRF (Bi-Directional Reflectivity Function) (Fig. 1(a)). Generally we define a BDRF as $R(\lambda,\phi i,\theta i,\phi v,\theta v)$ that relates incoming light in the direction (ϕ_i,θ_i) to outgoing light in the direction $(\phi v,\theta v)$. The BDRF is the ratio of outgoing intensity to incoming energy:

$$R(\lambda,\phi i,\theta i,\phi v,\theta v)=\frac{Iv(\lambda,\phi i,\theta i,\phi v,\theta v)}{Ei(\phi i,\theta i)} \quad (1)$$

Where the relationship between the incoming energy and incoming intensity is given by

$$Ei(\phi i, \theta i) = Ii(\phi i, \theta i)\cos(\theta i) \tag{2}$$



(a) (b)

Figure 1 – (a) Bi-Directional Reflectivity Function Model, (b) Camera viewpoint for feature detection.

In Fig. 1(b) a view of the camera position and strong shadow pixels is shown. In strong shadow pixels due to the lack of light source and low value of incoming energy, there is not considerable amount of outgoing intensity. Therefore pixels of shadow in under-vehicle region have the lowest intensity among image pixels. We demonstrated this feature of shadow pixels using their photometric properties [25], [26]. In the previous work [15], it was showed that each of the three RGB color components, if positive and not zero, decreases when passing from a lit region to a shadowed one, that is $R_{shadow} < R_{lit}$, $G_{shadow} < G_{lit}$, and $B_{shadow} < B_{lit}$. So this region has different spectral properties. Also other shadows in a traffic scene have surface reflection. They also have more intensity than under vehicle shadows.

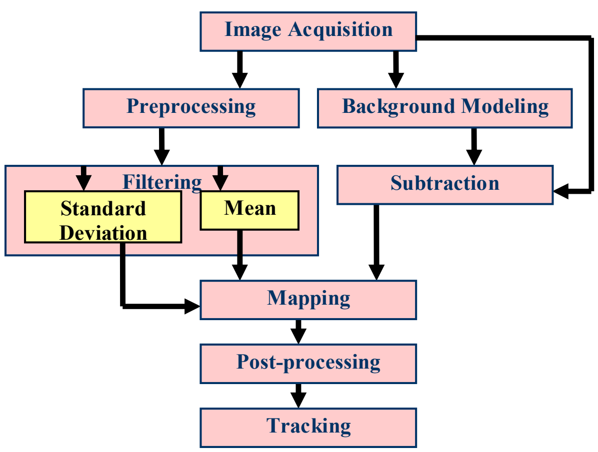


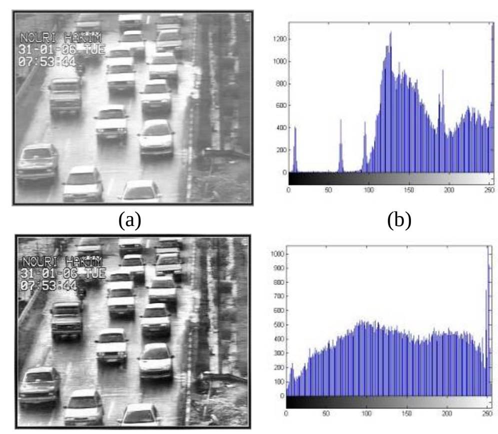
Figure 2 – Diagram of our approach.

III. Our Proposed Approach

The focus of this work is on the problem of strong shadow segmentation for on-road vehicle detection and tracking in the presence of occlusion. Strong shadow feature was initially presented in [15]. The detection of shadows was done by converting to gray level with different color-maps, thresholding the image, and applying some morphological operations. Stated succinctly, the shadow underneath a vehicle is darker than any other point on the road. We have proposed a new method for segmentation of strong shadow pixels, not having the previous problems of vehicle detection invariant to weather/lighting conditions on both wet and dry roads. Since underneath strong shadows have various sizes in different regions of the image, due to depth of perspective images, to avoid data loss in far regions, accurate parameters need to be set. Determining these critical values was the main problem in [15]. To solve this problem we have presented a new method that uses local mean and standard deviation of the regions around pixels to segment strong shadow. This section presents the proposed vehicle detection system, illustrated in Fig. 2.

A. Pre-processing

We have shown that this new method of vehicle detection can be significantly improved by means of simple content-adapted techniques. These techniques are brightness and contrast improvement according to the contents relevance when necessary. Adverse weather condition causes low contrast for whole pixels of the image like strong shadow pixels. The contrast was enhanced transforming the values using Contrast-Limited Adaptive Histogram Equalization (CLAHE). CLAHE operates on small regions in the image, called tiles, rather than the entire image (Fig. 3.).



(a) (b) (c) (d)

Figure 3 – (a) Real traffic scene in the rainy weather condition. (b) Histogram of original image. (c) Enhanced image. (d) Histogram of enhanced image.

B. Background Modeling and Subtraction

In [15] road was used to be masked in each frame image to determine the processing region. This attempt was to avoid the probable other shadow like areas in the image. But here we model the background and extract the moving objects. The strong shadow underneath the vehicle is moving frame to frame, like the vehicle itself. Doing this we omit all other dark areas in the image.

We have used the background modeling proposed by W^4, a real-time visual surveillance system [27]. In order to distinguish moving pixels from stationary pixels, First we apply a pixel wise median filter over time to several portions of video (In this experiment 100 frames were used, about 3.3 seconds). Only the stationary pixels are chosen as background pixels. Let A be a video sequence containing N consecutive frames, $A^k(i, j)$ be the intensity of pixel (i, j) in

k^{th} frame, $\sigma(i, j)$ and $\mu(i, j)$ be the standard deviation and median values at pixel (i, j) in all frames in video sequence A respectively. Background model $B(i, j) = [m(i, j), n(i, j), d(i, j)]$ ($m(i, j)$ minimum, $n(i, j)$ maximum intensity values and $d(i, j)$ maximum intensity difference between frames observed during the training period) is calculated using formula 3[27]:

$$B(i, j) = \begin{bmatrix} m(i, j) \\ n(i, j) \\ d(i, j) \end{bmatrix} = \begin{bmatrix} \min_z A^z(i, j) \\ \max_z A^z(i, j) \\ \max_z \left| A^z(i, j) - A^{z-1}(i, j) \right| \end{bmatrix} \quad (3)$$

where z, are frames satisfying:

$$\left| A^z(i, j) - \mu(i, j) \leq 2\sigma(i, j) \right| \quad (4)$$

Only stationary pixels are selected as background pixels. This is because when a moving object moves across a pixel, the intensity of that pixel decreases or increases sharply. Then if we choose the median value of that pixel over time, we can model image without any moving object, which is the background image we are looking for. After the training phase initial background model for each pixel is obtained (i.e. $B(i, j)$). The result is the background image and can be used in the background subtraction process.

For each frame, background subtraction yields to an image containing only the moving objects in that scene. After subtraction we convert the resulting image to a binary image. Fig. 4 illustrates an example of this background modeling approach. The result of the background subtraction phase is a binary image 'S':

$$S(i, j) = \begin{cases} 1 & A(i, j) \text{ part of the dark region} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

(a) (b) (c) (d)

Figure 4 – (a) Extracted background image. (b) A sample frame of the video sequence. (c) Background subtraction, subtracting frame image from the background. (d) Binary Image of c.

C. *Filtering and Mapping*

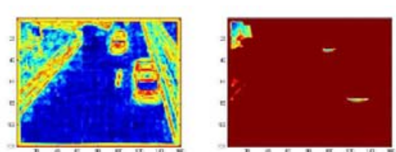
The intensity of the shadow pixels depends on the illumination of the image, which in turn depends on weather conditions. Therefore the thresholds are not by no means, fixed and it causes some implementation difficulties. Facing this problem we use multi-level processing to get accurate results in depth of perspective images. In this work we used the information acquired by the mean and standard deviation in the area around each image pixel to segment strong shadow pixels. Rainy weather conditions or bad illumination conditions make the color of the road pixels darker, but our results have shown satisfying outcome.

Local Standard Deviation (STD) and mean values in images have been widely used for pattern detection and image segmentation [28], [29]. Here we use these two parameters to detect strong shadow under each vehicle. A sliding square window is used as the neighborhood element. The length of this window is defined depending on the image perspective and depth. Let the window length be N, the 2D Arithmetic Mean (μ) and the 2D Standard Deviation (σ) for the neighborhood around pixel (i, j) of image 'I' is calculated using:

$$\mu(i, j) = \frac{1}{N^2} \sum_{k=\left\lceil -\frac{N}{2} \right\rceil}^{\left\lceil \frac{N}{2} \right\rceil} \sum_{l=\left\lceil -\frac{N}{2} \right\rceil}^{\left\lceil \frac{N}{2} \right\rceil} I(i+k, j+l) \quad (6)$$

$$\sigma(i, j) = \sqrt{\frac{1}{N^2 - 1} \sum_{k=\left\lceil -\frac{N}{2} \right\rceil}^{\left\lceil \frac{N}{2} \right\rceil} \sum_{l=\left\lceil -\frac{N}{2} \right\rceil}^{\left\lceil \frac{N}{2} \right\rceil} \left(I(i+k, j+l) - \mu(i, j) \right)^2} \quad (7)$$

Considering the formulas above and the properties of arithmetic mean and standard deviation, it's obvious that the mean filter smoothes the image regarding the pixels' intensities, and the standard deviation helps finding texture information in an image. $\sigma(i, j)$, in any solid texture area, has a very low value, since the variation of the intensities in that area is not too much. In such areas, that the intensities are not largely variable, $\mu(i, j)$ will be most like the pixels' values in that area. So, in the areas of strong shadow, regions of our interest, as far as dark areas have low intensities and they have dark solid textures $\mu(i, j)$ and $\sigma(i, j)$ are supposed to be generally low. Pixels with low values of both $\mu(i, j)$ and $\sigma(i, j)$ are likely to be parts of the strong shadow region. They are marked as candidates. Figure 5 illustrates the $\mu(i, j)$ and $\sigma(i, j)$ scaled images of the frame showed in Fig. 4(b).



(*a*) (*b*)

Figure 5 – (a) Scaled image of STDs. (b) Image of Means.

As discussed, the neighborhood window length needs to be adapted regarding the image depth. Since far vehicles in a perspective have smaller shadows. Treating the same with the vehicles far or near in a traffic scene yields to losing information. Therefore, we assigned three different levels in a processing image. The vehicles close to the camera are processed with a big window, the ones far in depth in the perspective are processed using a small window, and finally for the ones in the middle a not-so-big not-so-small window is used.

The next phase includes integrating results of all previous steps. This phase is called mapping. First, we threshold $\mu(i, j)$ and $\sigma(i, j)$ matrices and map the results on the image acquired by background subtraction, $S(i, j)$. So, all the pixels having the following condition are dark parts of the moving vehicle and are very likely to be the strong shadow:

$$\begin{aligned} &\mu(i, j) > mean\_threshold \\ &\sigma(i, j) > std\_threshold \\ &S(i, j) = 1 \end{aligned} \tag{8}$$

Fig. 6 shows the thresholded images of $\mu(i, j)$ and $\sigma(i, j)$ of Fig. 4(b) and the result of mapping using condition 8.

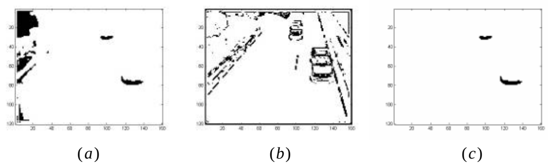


Figure 6 – (a) Thresholded mean. (b) Thresholded standard deviation. (c) Result of the mapping phase.

D. Post-processing and Localization

After the mapping process, we should count each individual component in the binary image as a symbol of a vehicle. In order to remove non-feature blobs, if the number of pixels in a component is less than a threshold we will remove it. This is done using some morphological techniques. Due to the different size of shadows, in this phase we used multilevel processing, as well. Therefore, we divided a traffic image to three different levels and determined three different values for the structuring element used in the morphological operation. Otherwise, using a fixed value for morphological operations might cause losing information of small, far shadows or counting non-feature shadows as vehicles. After this step, we count blobs as the representative of a vehicle.

E. Tracking

After extracting the strong shadow regions, we are provided with an image containing several blobs in. In this section we discuss how to track vehicles appropriately, even with a large number of occlusions, using this feature. As discussed, this feature had a likely low-cost detection process; so we use an easy approach for tracking in order not to demolish the low overall implementation and deployment cost.

The process of tracking blobs is essentially establishing correspondences between the blobs in consecutive frames in a video sequence. Our approach is based on the area each shadow region occupies. In consecutive frames we match the detected regions. In order to mark two blobs of a same shadow region in consecutive frames, they need to have overlapping pixels. This is formulated as what follows.

Let A_i^n be the area of the i[th] region in the n[th] frame, here it is the total number of pixels in that region. The overlapping area of a blob O_i^n in consecutive frames is the number of pixels common to both the regions:

$$O_i^n = A_i^n \cap A_i^{n-1} \tag{9}$$

For each frame we maintain an association table describing the association of each region in the current frame with the region in the previous frame regarding the overlap area calculated above. If a new region is detected in the current frame then a field is added in the table with a zero association while if a region exits the tracking area the corresponding field in the table is deleted.

F. Occlusion Handling

In traditional vehicle tracking techniques, occlusion has always been a considerable, challenging problem. Here with the proposed feature, occlusion problem is mostly solved. Theoretically, when the vehicles in a scene occlude, the strong shadows under them never merge, the occluding vehicles bodies prevent merging shadows underneath each. But in some rare cases (e.g. accidents or noisy not properly illuminated traffic scenes), we need to consider a probability for that to happen, for the sake of the reliability and stability of the system. As it will not be continuing for a long time and the two strong shadow regions will be separated again in the next coming frames, we need to temporarily solve this problem. To label the overlapping blobs, regarding the area of each in the previous frames, we share the pixels between them. As an example if two regions are merged and had separate areas A_1 and A_2 in the previous frame, the area will be shared to the two regions with the following ratios:

$$\frac{A_1}{A_1 + A_2}, \frac{A_2}{A_1 + A_2} \tag{10}$$

The first ratio is the share of the first blob from the merged region, and the second one is the share of the second blob.

The only problem could be that the strong shadow region of one vehicle be completely occluded with the body of another vehicle. In this case the first vehicle is lost. In our proposed model, when the area of the region is reduced in size less than a specified value, or has no overlapping area with any of the blobs in the proceeding frame, it is lost. It will be omitted from association table of that frame, and be pushed in the back of a queue together with a timestamp (e.g. the frame number). This queue indicates lost vehicles. Whenever a new blob is found in the intermediate pixels of a frame, not in the boundary pixels, one of the lost vehicles is again back in the image. In such a case, it will be matched with one of the blobs in the queue regarding its size and direction. The blob is popped from the queue and put back in the association table. If an element in the queue is not found back in the video for a long time, it is popped from the queue in order not to interfere in future processing.

IV. Experimental Results

Experimental results and comparisons using real data demonstrate the superiority of the proposed approach which has achieved an average accuracy of 94% on completely novel test images. Our tests have demonstrated the effectiveness of our approach in handling occlusion. The first image in Fig. 7, shows real traffic scene with considerable occlusion, whereas the second one in Fig. 8. illustrates a group of occluded vehicles in congested traffic and rainy weather which are detected accurately and finally Figure 9 shows result of our proposed approach in bad illumination and remarkable cast shadow condition. Our approach could ignore cast shadows precisely.

V. Conclusion

In this work, we used some statistical parameters of the pixels in each frame to detect the strong shadow under the vehicle. After enhancing the image quality, the background was extracted, for each frame background subtraction and image filtering was done. Mean and Standard Deviation matrixes together with the output of the background subtraction phase, are fed into a mapping process to extract the strong shadow regions. The post-processing phase helped leaving out the noise and non-desirable regions and finally, blobs were tracked.

We tested our approach on different traffic scenes, adverse weather conditions and noisy or cluttered images, and it showed accurate and considerable results. It also can ignore cast shadows on the street. Furthermore, it's very low-cost and easy to implement.

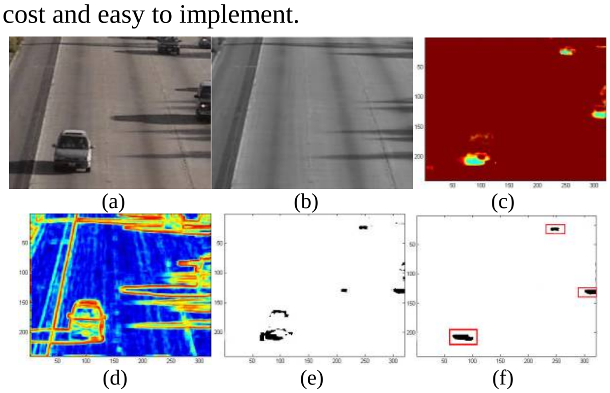


Figure 7 – (a) Original image containing cast shadows. (b) Background. (c) Result of Mean filter. (d) Result of the mapping phase. (e) Result of post-processing and localization.

VI. Future Works

Our focus in the future work is proposing optimal method for local enhancing regarding to weather and illumination conditions to make the algorithms more robust. The current results are outstanding and further work is being done to make the system more practical by combination with other approaches for better tracking results. Cases in which the strong shadows underneath the vehicles are lost, other features can be taken in use for reliable tracking or detection.

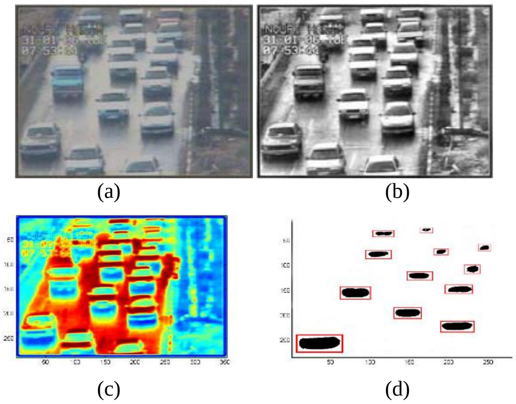


Figure 8 – (a) Real traffic scene in the rainy weather condition , (b) Enhanced image using histogram (CLAHE) (c) Result of Mean filter (d) Detected vehicles

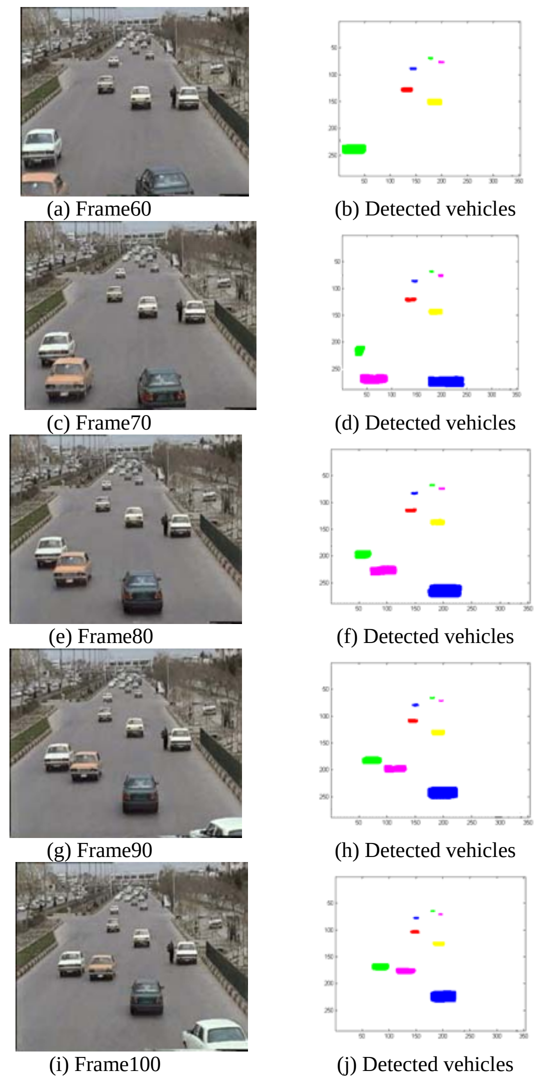


Figure 9 – Tracked vehicles in frames 60,70,80,90,100.

REFERENCES

[1] S. Gupte, O. Masoud, R. F. K. Martin, and N. P. Papanikolopoulos. "Detection and classification of vehicles". *IEEE Transactions on Intelligent Transportation Systems*, 3(1):37–47, March 2002.

[2] D. Magee. "Tracking multiple vehicles using foreground, background and motion models". *ECCV Workshop on Statistical Methods in Video Processing,* 2002.

[3] D. Daily, F.W. Cathy, and S. Pumrin. "An algorithm to estimate mean traffic speed using uncalibrated cameras". *IEEE Conf. for ITS*, pages 98–107, 2000.

[4] D. Koller, J. Weber, and J. Malik. "Robust multiple car tracking with occlusion reasoning". In *European Conference on Computer Vision*, pages 189–196, 1994.

[5] D. Koller, K Dandilis, and H. H. Nagel. "Model based object tracking in monocular image sequences of road traffic scenes". *International Journal of Computer Vision*, 10(3):257– 281, 1993.

[6] M. Haag and H. Nagel. "Combination of edge element and optical flow estimate for 3Dmodel vehicle tracking in traffic image sequences". *International Journal of Computer Vision,*295–319, Dec 1999.

[7] J. M. Ferryman, A. D. Worrall, and S. J. Maybank. "Learning enhanced 3d models for vehicle tracking". In *British Machine Vision Conf.*, pages 873–882, 1998.

[8] C. Schlosser, J. Reitberger, and S. Hinz. "Automatic car detection in high resolution urban scenes based on an adaptive 3D-model". In *EEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, Berlin*, pages 98–107, 2003.

[9] S. Kamijo, K. Ikeuchi, and M. Sakauchi. "Vehicle tracking in low-angle and front view images based on spatio-temporal markov random fields". *8th World Congress on Intelligent Transportation Systems*, 2001.

[10] D. Beymer, P. McLauchlan, B. Coifman, and J. Malik. A real time computer vision system for measuring traffic parameters. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 495–501, 1997.

[11] Jitendra Malik and Stuart Russel. "Final Report for Traffic Surveillance and Detection Technology Development New Traffic Sensor Technology", Dec. 1996.

[12] Chachich, A. A. Pau, A. Barber, K. Kennedy, E. Oleiniczak, J. Hackney, Q. Sun, E. Mireles, "Traffic sensor using a color vision method," *Proceedings of the International Society for Optical Engineering*, Vol. 2902, January, 1997, pp. 156-164.

[13] Zehang Sun, George Bebis and Ronald Miller. "Improving the Performance of On-Road Vehicle Detection by Combining Gabor and Wavelet Features". *The 2002 IEEE International Conference on Intelligent Transportation* Systems, September 2002.

[14] Kai She; Bebis, G.; Haisong Gu; Miller, R., "Vehicle tracking using on-line fusion of color and shape features". The 7th International IEEE Conference on Intelligent Transportation Systems Oct. 2004.

[15] M. Sadeghi, M. Fathy. "A Low-cost Occlusion Handling Using a Novel Feature In Congested Traffic Images". In Proc. of IEEE ITSC 2006. Toronto, 2006, pp. 522-527.

[16] A. Huertas, R. Nevatia. "Detecting buildings in aerial images". Comput. Vis. Graph. Image Process. Vol. 41, 1988, pp. 31–152.

[17] A. Yoneyama, C.H. Yeh, C. Kuo. "Moving cast shadow elimination for robust vehicle extraction based on 2d joint vehicle/shadow models". IEEE Conf. on Advanced Video and Signal Based Surveillance, July 2003, Miami, USA.

[18] J.M. Scanlan, D.M. Chabries, R. Christiansen. "A shadow detection and removal algorithm for 2-d images". In Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, 1990, pp. 2057–2060.

[19] M. Adjouadj. "Image analysis of shadows, depressions, and upright objects in the interpretation of real world scenes". In Proc. IEEE Int. Conf. on Pattern Recognition (ICPR), 1986, pp. 834– 838.

[20] G.S.K. Fung, N.H.C. Yung, G.K.H. Pang, A.H.S. Lai. Effective moving cast shadows detection for monocular color image sequence. In Proc. 11th Int. Conf. on Image Analysis and Processing, 2001, pp. 404–409.

[21] S. Nadimi, B. Bhanu. Moving shadow detection using a physicsbased approach. in Proc. IEEE Int. Conf. Pattern Recognition, vol. 2, 2002, pp. 701–704.

[22] R. Gershon, A.D. Jepson, J.K. Tsotsos. "Ambient illumination and the determination of material changes". Journal of the Optical Society of America A, VOL.3, NO.10, October 1986, pp. 1700-1707.

[23] A. Cavallaro, E. Salvador and T. Ebrahimi. "Shadow-aware object-based video processing". IEE Proc.-Vis. Image Signal Process., Vol. 152, No. 4, August 2005, pp. 398-406.

[24] T. Gevers, H. Stokman. "Classifying color edges in video into shadow-geometry, highlight, or material transitions", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 5, NO. 2, JUNE 2003, pp. 237-243.

[25] S.A. Shafer. "Using color to separate reflection components". Color Res. Appl. (10)-4, 1985, 210–218.

[26] D.A. Forsyth, J. Ponce. :Computer Vision: A Modern Approach". Prentice Hall, New York, 2003.

[27] I. Haritaoglu, D. Harwood, and L. S. David. "W^4: Real-time Surveillance of People and Their Activities". IEEE Transaction on Pattern Recognition and Machine Intelligence, 22(8):809-830, 2000.

[28] C. Tornieri, F. Bremond, M. Thonnat. Updating of the reference Image for Visual Surveillance Systems. Intelligent Distributed Surveillance Systems - IDSS 2003, London, UK, Feb. 2003.

[29] C. Wolf, J.-M. Jolion, and F. Chassaing. Text localization, enhancement and binarization in multimedia documents. In Proceedings of the ICPR 2002, volume 2, pages 1037-1040, August 2002.