A New Approach for Vehicle Detection in Congested Traffic Scenes Based on Strong Shadow Segmentation

Ehsan Adeli Mosabbeb[1], Maryam Sadeghi[2], and Mahmoud Fathy[1]

[1] Computer Eng. Department, Iran University of Science and Technology, Tehran, Iran
[2] Computer Science Department, Simon Fraser University, Vancouver, Canada
eadeli@comp.iust.ac.ir, msa68@sfu.ca, mahfathy@iust.ac.ir

Abstract. Intelligent traffic surveillance systems are assuming an increasingly important role in highway monitoring and city road management systems. Recently a novel feature was proposed to improve the accuracy of object localization and occlusion handling. It was constructed on the basis of the strong shadow under the vehicle in real-world traffic scene. In this paper, we use some statistical parameters of each frame 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 cast shadows, bad illumination conditions and occlusions, taken from both outdoor highways and city roads are presented.

1 Introduction

Increasing congestion on freeways has generated an interest in new vehicle detection technologies such as video image processing. Existing commercial image processing systems work well in free-flowing traffic, but the systems have problems with congestion, occlusion, shadows and lighting transitions. This paper addresses the problem of vehicle segmentation in traffic images including vehicle occlusion and cast shadows.

Some of the related works for analyzing surveillance images are based on background subtraction methods [1, 2], and some use an extended Kalman filter [3, 4]. The early attempts to solve the occlusion problem involved simple thresholding, while later methods applied energy minimization and motion information [5, 6]. More recently, motion segmentation methods based on active contours [7] have been proposed. The other concept of object tracking as spatio-temporal boundary detection has been proposed in [8]. The advantage of part-based methods is shown in [9], and the algorithm known as predictive Trajectory Merge-and-Split (PTMS) in [10], has been developed to detect partial or complete occlusions during object motion. In [11] 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.

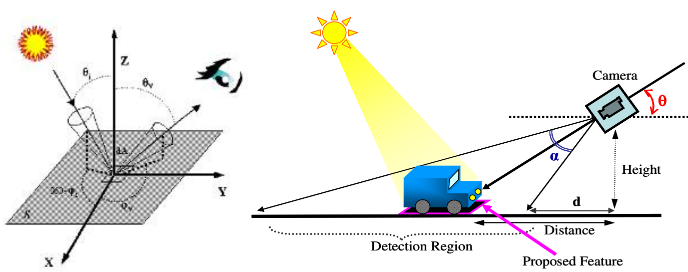
The problem of shadow detection has been increasingly addressed over the past years. Shadow detection techniques can be classified into two groups: model-based and property-based techniques. Model-based techniques are designed for specific

G. Bebis et al. (Eds.): ISVC 2007, Part II, LNCS 4842, pp. 427–436, 2007.

applications, such as aerial image understanding [12] and video surveillance [13]. Luminance information is exploited in early techniques by analyzing edges [14], and texture information [15]. Luminance, chrominance and gradient density information is used in [16]. Color information is used also in [17]. A physics-based approach to distinguish material changes from shadow boundaries in chromatic still images is presented in [18]. Cavallaro et. al. in [19] 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 [20]. Using strong shadow information as a feature for vehicle detection was initially discussed in [11]. In this first attempt, it was found that the area under a vehicle is distinctly darker than any other areas on an asphalt paved road.

2 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. Geometric properties need a priori information such as object size or direction of light rays. We can model geometric properties of shadow and illumination with BDRF (Bi-Directional Reflectivity Function) (Figure 1(a)).



(a) (b)

Fig. 1. (a) BDRF Model, (b) Camera viewpoint for feature detection

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 (ϕ_v,θ_v). The BDRF is the ratio of outgoing intensity to incoming energy:

$$R(\lambda,\phi_i,\theta_i,\phi_v,\theta_v) = \frac{Iv(\lambda,\phi_i,\theta_i,\phi_v,\theta_v)}{Ei(\phi_i,\theta_i)} \tag{1}$$

where the relationship between the incoming energy and incoming intensity is

$$E_i(\phi_i,\theta_i) = I_i(\phi_i,\theta_i)\cos(\theta_i) \tag{2}$$

In Figure 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 can demonstrate this feature using the photometric properties [13, 21].

$$L_r(\lambda,p) = L_a(\lambda) + L_b(\lambda,p) + L_s(\lambda,p) \tag{3}$$

Where $L_a(\lambda)$, $L_b(\lambda, p)$, $L_s(\lambda, p)$ are the ambient reflection term, the body reflection term, and the surface reflection term, respectively and λ is the wavelength.

$$L_r(s\mathrm{hadow})(\lambda,p) = L_a(\lambda) \tag{4}$$

$$C_i(x,y) = \int E(\lambda, x, y)\, Sci(\lambda, x, y) d\lambda \tag{5}$$

$$Sci(\lambda) \in \{SR(\lambda), SG(\lambda), SB(\lambda)\} \tag{6}$$

$$C_i(x,y)\mathrm{lit} = \int \alpha(L_a(\lambda) + L_b(\lambda, p) + L_s(\lambda, p))Sci(\lambda)d\lambda \tag{7}$$

$$C_i(x,y)\mathrm{shadow} = \int \alpha(L_a(\lambda))Sci(\lambda,x,y)d\lambda \tag{8}$$

In which $C(x, y)_{shadow} = (R_{shadow}, G_{shadow}, B_{shadow})$. It follows 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

$$R_{shadow} < R_{lit},\ G_{shadow} < G_{lit},\ B_{shadow} < B_{lit} \tag{9}$$

So this region has different spectral properties. Also other shadows in traffic scene have $L_s(\lambda ,p)$ and more intensity than under vehicle shadows.

3 Our Proposed Approach

The focus of this work is on the problem of strong shadow segmentation for on-road vehicle detection and occlusion handling. Strong shadow feature was initially presented in [11]. There, contrary to previous algorithms which had been implemented for shadow removal to detect foreground objects, shadow was used as a useful feature to detect vehicles. The detection of shadows was done by converting to gray level with different colormaps, thresholding the image, and applying some morphologic operations. 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. Cast shadows are effectively removed, while strong shadows remain. 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 of images, accurate parameters need to be set. Determining these critical values was the main problem in [11]. To solve this problem we have presented a new method, that uses mean and standard deviation information. This section presents the proposed vehicle detection system, illustrated in Figure 2.

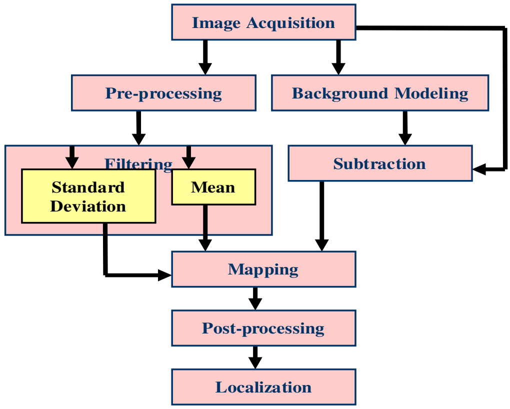
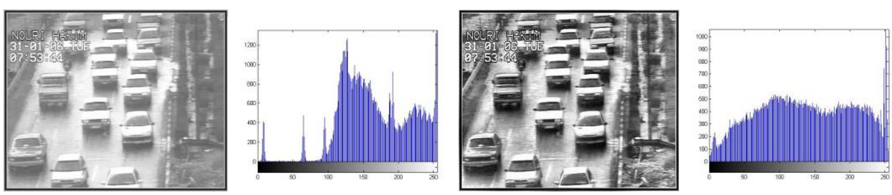


Fig. 2. Diagram of our approach

3.1 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 all of the pixels in an image. Because of differences in contrast and brightness in different weather conditions the contrast of the intensity image were enhanced by 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 (Figure 3).



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

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

3.2 Background Modeling and Subtraction

Sadeghi and Fathy in [11] used to mask the road 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 from frame to frame, like the vehicle itself. Doing this we omit all other probable static dark areas in the image.

The background modeling used here is proposed by W^4, a real-time visual surveillance system [22]. 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, 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 kth frame, $\sigma(i,j)$ *and* $\mu(i,j)$ be the standard deviation and median values at pixel (i,j) in all frames in 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 as [22]:

$$B(i,j)=\begin{bmatrix} m(i,j) \\ n(i,j) \\ d(i,j) \end{bmatrix}=\begin{bmatrix} \min_z A^z(i,j) \\ \max_z A^z(i,j) \\ \max_z \left|A^z(i,j)-A^{z-1}(i,j)\right| \end{bmatrix} \tag{10}$$

where z, are frames satisfying:

$$\left|A^z(i,j)-\mu(i,j)\le 2\sigma(i,j)\right| \tag{11}$$

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. $B(i,j)$). We convert each frame to gray-scale, feeding into the W^4 algorithm. The result is the background and can be used in the background subtraction process.

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

Fig. 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.

For each frame, background subtraction yields to an image containing only the moving objects in that scene. As far as we are seeking for the strong shadow under the vehicle, we subtract the frame image from the background, truncating out of range pixel values. In a gray-scale image, the values of the strong shadow pixels, being very dark, are likely the lowest. But the same pixels in the background have values larger than those shadow pixels. Subtracting the frame image from the background image would make the darker data outstanding. The strong shadow is a part of these dark pixels. After subtraction we convert the resulting image to a binary image. Figure 4 illustrates an example of this background modeling approach. The result of the background subtraction phase is a binary image ‘S’:

$$S(i,j)=\begin{cases}1 & A(i,j)\ \text{part of the dark region}\\ 0 & \text{otherwise}\end{cases} \tag{12}$$

3.3 Filtering and Mapping

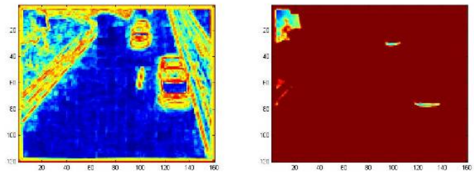
In [11] authors have used converting images to gray level with different color maps, thresholding the image, and applying some morphological processing. The intensity of the shadow pixels depends on the illumination of the image, which in turn depends on weather conditions. Therefore the thresholds are not fixed causing implementation difficulties. Different threshold values used for image to binary conversion was a problem. Facing this problem, multi-level processing is used 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 [23]. 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:

$$\mu(i,j)=\frac{1}{N^2}\sum_{k=\left\lceil -\frac{N}{2}\right\rceil}^{\left\lfloor \frac{N}{2}\right\rfloor}\sum_{l=\left\lceil -\frac{N}{2}\right\rceil}^{\left\lfloor \frac{N}{2}\right\rfloor} I(i+k,j+l) \tag{13}$$

$$\sigma(i,j)=\sqrt{\frac{1}{N^2-1}\sum_{k=\left\lceil -\frac{N}{2}\right\rceil}^{\left\lfloor \frac{N}{2}\right\rfloor}\sum_{l=\left\lceil -\frac{N}{2}\right\rceil}^{\left\lfloor \frac{N}{2}\right\rfloor}\left(I(i+k,j+l)-\mu(i,j)\right)^2} \tag{14}$$

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 6 illustrates the $\mu(i,j)$ and $\sigma(i,j)$ scaled images of the frame showed in 4(b).



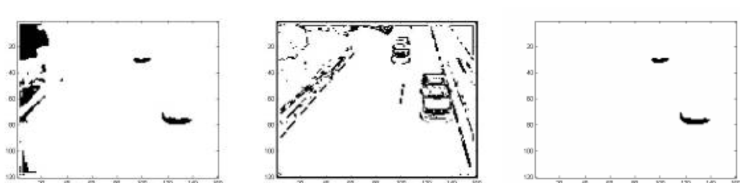
(a) (b)

Fig. 5. (a) Image of Means. (b) Scaled image of Standard Deviations.

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 some 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:

$$\begin{aligned} &\mu(i,j) > mean\_threshold \\ &\sigma(i,j) > std\_threshold \\ &S(i,j) = 1 \end{aligned} \tag{15}$$



(a) (b) (c)

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

Figure 7 shows the thresholded images of $\mu(i,j)$ and $\sigma(i,j)$ of figure 4(b) and the result of mapping using 15.

3.4 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-desired blobs, erosion and dilation morphologic techniques are taken into account. Due to different sizes of shadows, multilevel processing is used, as well. A traffic image is divided into three different levels; three different values for the structuring element are determined to be used in the morphologic operation. Using a fixed value for morphologic 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.

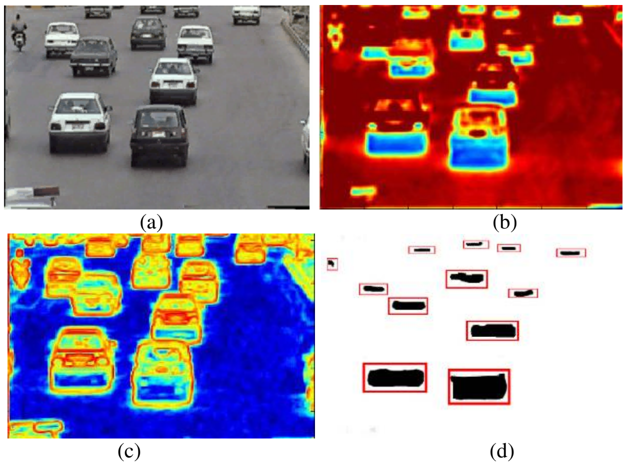


Fig. 7. (a) Occluded traffic scene. (b) Result of Mean filter. (c) Result of STD filter. (d) Detected vehicles.

4 Experimental Results

Experimental results and comparisons using real data demonstrate the superiority of the proposed approach which has achieved an average accuracy of 93.94% on completely novel test images. Our tests have demonstrated the effectiveness of our approach handling occlusion in real traffic images including high congestion, noisy, cluttered, snowy, and rainy and scenes containing remarkable shadows, bad illumination conditions and occlusions.

The first image in Figure 8 shows real traffic scene with considerable occlusion, whereas the second one in Figure 9 illustrates a group of occluded vehicles in congested

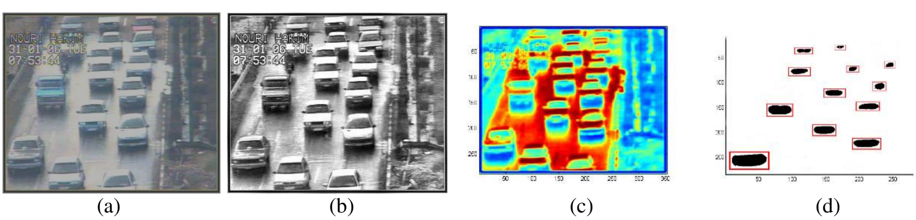


Fig. 8. (a) Real traffic scene in the rainy weather condition , (b) Enhanced image using histogram (CLAHE) (c) Result of Mean filter (d) Detected vehicles

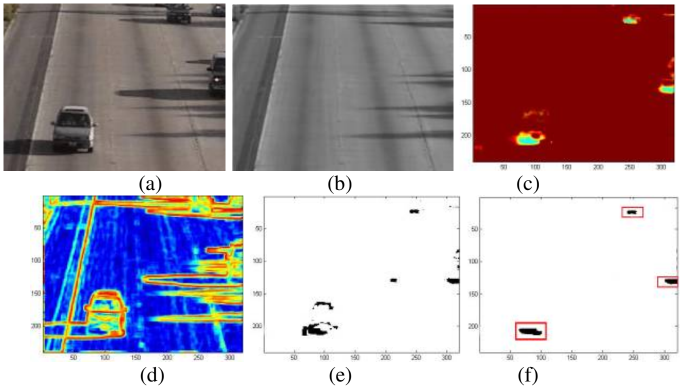


Fig. 9. (a) Original image containing cast shadows. (b) Background. (c) Result of Mean filter. (d) Result of the mapping phase. (e) Result of post-processing and localization.

traffic and rainy weather which are detected accurately and finally Figure 10 shows result of our proposed approach in bad illumination and remarkable cast shadow condition. Our approach could ignore cast shadows precisely.

5 Conclusion and Future Works

In this work we first reviewed a recently proposed feature for vehicle pose detection, the strong shadow under any vehicle. In order to detect this region in the video sequence 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.

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

Our focus in future work is proposing optimal method for local enhancing regarding to weather and illumination conditions to make the algorithms more robust. Also we are working to convey Mean and STD filters on results of subtraction phase and moving object detection to improve the time-cost. The current results are outstanding and further work is being done to make the system more practical.

References

1. Gutchess, D., Trajkovics, M., Cohen-Solal: A background model initialization algorithm for video surveillance. In: Proc. of IEEE ICCV 2001, Pt.1, pp. 744-740 (2001)
2. Javed, O., Shafique, K., Shah, M.: A hierarchical approach to robust background subtraction using color and gradient information. In: Workshop on Motion and Video Comp. pp. 22–27 (2002)

3. Veeraraghavan, H., Masoud, O., Papanikolopoulos, N.: Computer vision algorithms for intersection monitoring. IEEE Trans. Intell.Transport. Syst. 4, 78–89 (2003)
4. Jung, Y., Lee, K., Ho., Y.: Content-Based event retrieval using semantic scene interpretation for automated traffic surveillance. IEEE Transaction ITS 2, 151–163 (2001)
5. Memin, E., Perez, P.: Dense estimation and object-based segmentation of the optical flow with robust techniques. IEEE Trans. Image Process 7(5), 703–719 (1998)
6. Chang, M., Tekalp, A., Sezan, M.: Simultaneous motion estimation and segmentation. IEEE Trans.Image Process 6, 1326–1333 (1997)
7. Ristivojević, M., Konrad, J.: Joint space-time motion-based video segmentation and occlusion detection using multiphase level sets. In: IS&T/SPIE Symposium on Electronic Imaging, Visual Communications and Image Processing, San Jose, CA, USA, pp. 18–22 (2004)
8. Mitiche, A., El-Feghali, R., Mansouri, A.-R.: Tracking moving objects as spatio-temporal boundary detection. In: IEEE Southwest Symp. on Image Anal. Interp., pp. 110–206 (April 2002)
9. Nowak, E., Jurie, F.: Vehicle categorization: Parts for speed and accuracy. UJF – INPG, Societe Bertin - Technologies, Aix-en-Provence (2005)
10. Melo, J., Naftel, A., Bernardino, A., Santos-Victor, J.: Viewpoint independent detection of vehicle trajectories and lane geometry from uncalibrated Traffic Surveillance Cameras. In: ICIAR Conf. on Image Analysis and Recognition, Porto,Portugal, September 29-October 1 (2004)
11. Sadeghi, M., Fathy, M.: A Low-cost Occlusion Handling Using a Novel Featur. In: Congested Traffic Images. In: proceeding of IEEE ITSC 2006 Toronto pp. 522–527 (2006)
12. Huertas, A., Nevatia, R.: Detecting buildings in aerial images. Comput. Vis. Graph. Image Process 41, 31–152 (1988)
13. Yoneyama, A., Yeh, C.H., Kuo, C.: Moving cast shadow elimination for robust vehicle extraction based on 2d joint vehicle/shadow models. In: IEEE Conf. on Advanced Video and Signal Based Surveillance, Miami, USA (July 2003)
14. Scanlan, J.M., Chabries, D.M., Christiansen, R.: A shadow detection and removal algorithm for 2-d images. In: Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), pp. 2057–2060 (1990)
15. Adjouadj, M.: Image analysis of shadows, depressions, and upright objects in the interpretation of real world scenes. IEEE Int. Conf. on Pattern Recog (ICPR), pp. 834–838 (1986)
16. Fung, G.S.K., Yung, N.H.C., Pang, G.K.H., Lai, A.H.S.: Effective moving cast shadows detection for monocular color image sequence. In: Proc. 11th ICIAP, pp. 404–409 (2001)
17. Nadimi, S., Bhanu, B.: Moving shadow detection using a physicsbased approach. In: Proc. IEEE Int. Conf. Pattern Recognition, vol. 2, pp. 701–704 (2002)
18. Gershon, R., Jepson, A., Tsotsos, J.: Ambient illumination and the determination of material changes. Journal of the Optical Society of America A 3(10), 1700–1707 (1986)
19. Cavallaro, A., Salvador, E., Ebrahimi, T.: Shadow-aware object-based video processing. IEE Proc.-Vis. Image Signal Process 152(4), 398–406 (2005)
20. Gevers, T., Stokman, H.: Classifying color edges in video into shadow-geometry, highlight, or material transitions. IEEE Trans. on Multimedia 5(2), 237–243 (2003)
21. Forsyth, D., Ponce, J.: Computer Vision: A Modern Approach. Prentice-Hall, NY (2003)
22. Haritaoglu, I., Harwood, D., David, L.S.: W^4 Real-time Surveillance of People and Their Activities. IEEE Trans. on Pattern Recog. and Machine Intelligence 22(8), 809–830 (2000)
23. Wolf, C., Jolion, J.-M., Chassaing, F.: Text localization, enhancement and binarization in multimedia documents. In: Proc. of the ICPR 2002, vol. 2, pp. 1037–1040 (August 2002)