International Journal of Image Processing and Vision Science

Volume 1 | Issue 3 Article 5

January 2013

CORRECTING FALSE SEGMENTATION IN VIDEO USING IMAGE OVER-SEGMENTATION

SHAILAJA SURKUTLAWAR SURKUTLAWAR
EXTCVivekananda education society of information technology, Mumbai, India, shaili_ayush@yahoo.com

RAMESH K KULKARNI
EXTCVivekananda education society of information technology, Mumbai, India, rk1_2002@yahoo.com

Follow this and additional works at: https://www.interscience.in/ijipvs

Part of the Robotics Commons, Signal Processing Commons, and the Systems and Communications Commons

Recommended Citation
SURKUTLAWAR, SHAILAJA SURKUTLAWAR and KULKARNI, RAMESH K (2013) "CORRECTING FALSE SEGMENTATION IN VIDEO USING IMAGE OVER-SEGMENTATION," *International Journal of Image Processing and Vision Science*: Vol. 1 : Iss. 3 , Article 5.
DOI: 10.47893/IJIPVS.2013.1032
Available at: https://www.interscience.in/ijipvs/vol1/iss3/5

This Article is brought to you for free and open access by the Interscience Journals at Interscience Research Network. It has been accepted for inclusion in International Journal of Image Processing and Vision Science by an authorized editor of Interscience Research Network. For more information, please contact sritampatnaik@gmail.com.

CORRECTING FALSE SEGMENTATION IN VIDEO USING IMAGE OVER-SEGMENTATION

SHAILAJA SURKUTLAWAR[1] & RAMESH K KULKARNI[2]

[1,2]EXTCVivekananda education society of information technology, Mumbai, India
E-mail : shaili_ayush@yahoo.com, rk1_2002@yahoo.com

Abstract - Moving objects detection is a fundamental step in many vision based applications. Background subtraction is the typical method. When scene exhibits pertinent dynamism method based on mixture of Gaussians is a good balance between accuracy and complexity, but fails due to two kinds of false segmentations i.e moving shadows incorrectly detected as objects and some actual moving objects not detected as moving objects. In computer vision, segmentation refers to process of partitioning a digital image in to multiple segments and goal of segmentation is to simplify and/or change representation of image in to something that is more meaningful and easier to analyse. A colour clustering based on k-means and image over-segmentation are used to segment the input frame into patches and shadow suppression done by HSV colour space, the outputs of mixture of Gaussians are combined with the colour clustered regions to a module for area confidence measurement. In this way, two major segment errors can be corrected. Experimental results show that the proposed approach can significantly enhance segmentation results.

***Keywords**— Adaptive Mixture of Gaussian, K-means ,HSV colour space , image over-segmentation.*

I. INTRODUCTION

Moving Objects segmentation is a fundamental and critical task in many vision based applications, such as automated visual surveillance, human-machine interface, and very low-bandwidth telecommunications. A common approach is to perform background subtraction, which identifies moving objects from the difference between the current frame and a reference frame (which often called “background model”). The background model must be representation of the scene with no moving objects and must be kept regularly updated because for some cases, the background is changing when time passes by. Such as view captured by an outdoor surveillance camera, the background is different when sun-light or weather is different. With respect to the state of the art [1-3], a wide variety of approaches performing background subtraction have been developed. A good review for these methods can be found in [4]. Referring to the conclusions of [4], Mixture of Gaussians [5-7] and Kernel density estimation (KDE) [8] can model well the background pdf in general cases and provide higher accuracy compared to other reviewed methods. If speed is concerned, they both have a constant complexity. But KDE has a much higher memory requirement (in order of a 100 frames).

So in the real applications, Mixture of Gaussians is the most frequently used method, as witnessed by the huge amount of literature on it. However, this method also suffers from slow learning at the beginning [9], incapability of identifying moving shadows from the objects casting them [6,10] and unsatisfied results in some cases.The video of the background model is not a literal visualization but it's simply a weighted sum of all components, whether they're part of the background model or not. From the results we can find that due to low rate of background updating, moving shadows, and possible influence by noise, the performance is poor. Though efforts have been imposed by many researchers to improve the algorithm in different senses, we have to say that a comprehensive physical model of the background is really difficult to develop. Therefore, a good post processing may be more suitable in general cases. In this paper, a novel color clustering based post-processing method is proposed, and will be discussed in details in the following sections.

II. BACKGROUND SUBSTRACTION

In the model of Mixture of Gauss [5-7], the background is not a single frame without any moving objects. Gaussian Mixture Model (GMM) is thought to be one of the best background modeling methods and works well when gradual changes appear in the scene . The GMM method models the intensity of each pixel with a mixture of K Gaussian distributions. The probability that a certain pixel has a value X_t at time can be written as

$$P(X_t) = \sum_{k=1}^{k} \omega_{k,t} \cdot \eta(X_t, \mu_{k,t}, \Sigma_{k,t}) \qquad (1)$$

where K is the number of distributions (currently, from 3 to 5 is used), $\omega_{k,t}$ is the weight of the kth Gaussian in the mixture at time t , and $\eta(X_t, \mu_{k,t}, \Sigma_{k,t})$ is the Gaussian probability density function.

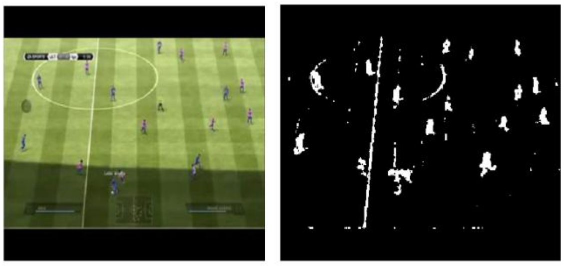
$$\eta(X_t, \mu_{k,t}, \Sigma_{k,t}) = \frac{1}{(2\pi)^{3/2} |\Sigma_{k,t}|^{1/2}} e^{\left\{\frac{-1(X_t-\mu_{k,t})^T \Sigma_{k,t}^{-1} (X_t-\mu_{k,t})}{2}\right\}}$$

where $\mu_{k,t}$ is the mean value and $\sum_{k,t}$ is the covariance of the k_{th} Gaussian at time *t*. For computational reasons, the covariance matrix is assumed to be of the form.

$$\sum_{k,t} = \sigma^2 . I \qquad (3)$$

Where σ is the standard deviation.

This assumes that the red, green, and blue pixel values are independent and have the same variance, allowing us to avoid a costly matrix inversion at the expense of some accuracy.



(a) (b)

Fig.1 (a) original frame (b) Extracted moving regions by mixture of Gaussians

III. MOVING SHADOW SUPPRESSION

Shadows are due to the occlusion of light source by an object in the scene. In particular, that part of the object not illuminated is called self-shadow, while the area projected on the scene by the object is called cast shadow [2]. This last one is more properly called moving cast shadow if the object is moving. In literature, many works have been published on shadow detection topic. Jiang and Ward [2] extract both self-shadows and cast shadows from a static image. They use a three level processes approach: the low level process extracts dark regions by thresholding input image; the middle level process detects features in dark regions, such as the vertexes and the gradient of the outline of the dark regions and uses them to further classify the region as penumbra (part of the shadow where the direct light is only partially blocked by the object), self-shadow or cast shadow; the high level process integrates these features and confirms the consistency along the light directions estimated from the lower levels.

Since our work addresses the problem of segmentation of moving objects, we aim to define an approach for detecting moving cast shadows on the background, without computing static shadows (due to static objects). In [3], the authors detail the shadow handling system using signal processing theory. Thus, the appearance of a point belonging to a cast shadow can be described as:

$$S_k(x, y) = E_k(x, y)\ \rho_k(x, y) \qquad (4)$$

where S_k is the image luminance of the point of coordinate (x,y) at time instant *t*. $E_k(x, y)$ is the irradiance and it is computed as follows:

$$E_k(x, y) = \begin{cases} C_A + C_P \cos\angle(N(x, y), L) & \textit{illuminate} \\ C_A & \textit{shadowed} \end{cases} \qquad (5)$$

where C_A and C_P are the intensity of the ambient light and of the light source, respectively, L the direction of the light source and $N(x,y)$ the object surface normal. $\rho_k(x, y)$ is the reflectance of the object surface.

In [3], some hypotheses on the environment are outlined:

I. strong light source
II. static background (and camera)
III. planar background

Most of the papers take implicitly into account these hypotheses. In fact, typically the first step computed for shadow detection is the difference between the current frame and a reference image, that can be the previous frame, as in [3], or a reference frame, typically named background model [4][5][6][1].

we can write this difference $D_k(x, y)$ as:

$$D_k(x,y) = S_{k+1}(x,y) - S_k(x,y) \qquad (6)$$

Let us consider that a previously illuminated point is covered by a cast shadow at frame $k + 1$. According to the hypothesis 2 of a static background, reflectance $\rho_k(x, y)$ of the background does not change with time, thus we can assume that

$$\rho_{k+1}(x,y) = \rho_k(x,y) = \rho(x, y) \qquad (7)$$

$$D_k(x, y) = \rho(x, y)\ C_P \cos\angle(N(x, y), L) \qquad (8)$$

Thus, if hypothesis 1 holds, C_p in eq.8 is high. Summarizing, if hypotheses 1 and 2 hold, difference in eq. 6 is high in presence of cast shadows covering a static background. This implies (as assumed in many papers) that shadow points can be obtained by thresholding the frame difference image. Eq. 8 detects not only shadows, but also foreground points. The papers in literature mainly differ in the way they distinguish between those points. In [4] Kilger uses a background suppression technique to find the moving objects and moving cast shadows in the scene. Then, for each object, it exploits the information on date, time and heading of the road computed by its system to choose whether to look for vertical or horizontal edges to separate shadows from objects. In [7], a the statistical a-posteriori estimation of the pixel probabilities of membership to the class of background, foreground or shadow points. The authors use three sources of information: local, based on the assumption that the appearance of a shadowed pixel can be approximated using a linear transformation of the underlying pixel appearance, according with the fact that the difference of eq. 8 should be positive; spatial, which iterates the local computation by re-computing the a-priori probabilities using the a-posteriori probabilities of the neighborhood; temporal, which predicts the position of shadows and objects from previous frames, therefore adapting the a-priori probabilities. The approach in [3] exploits the local appearance change due to shadow by computing the ratio $R_k(x, y)$ between the appearance of the pixel in the actual frame and the appearance in a reference frame.

$$R_k(x,y) = \frac{S_{k+1}(x,y)}{S_k(x,y)} \quad (9)$$

that can be rewritten as ratio between irradiance and reflectance by using eqs. 4 and 7 :

$$R_k(x,y) = \frac{E_{k+1}(x,y)}{E_k(x,y)} \quad (10)$$

If a static background point is covered by a shadow, we have:

$$R_k(x,y) = \frac{C_A}{C_A + C_p \cos \angle (N(x,y), L)} \quad (11)$$

This ratio is less than one. In fact, the angle between N(x, y) and L is in the range between $-\pi/2$ to $\pi/2$ therefore the Cos function is always positive. Moreover, due to hypothesis 3, we can assume N(x, y) as spatially constant in a neighbourhood of the point, because the background is supposed planar in a neighbourhood .In [3], authors exploit the spatial constancy of N to detect shadows by computing the variance in a neighbourhood of the pixel of the ratio R_k (x, y): a low variance means that assumption 3 holds, then they mark that pixel as "possible shadow". Moreover, authors use a lot of other techniques in order to exploit all the four assumptions (such as edge detection and gradient calculation). eq.11 can be seen as the ratio between the luminance after and before shadow appears. In a similar way, Davis et al. ([5][8]) define a local assumption on the ratio between shadow and shadowed point luminance. This is based on the hypothesis that shadows darken the covered point, as eq. 11 and the considerations above confirm. This approach has been improved in [6] where the authors state that shadow has similar chromaticity but lower brightness than that of the same pixel in the background image. They base this statement on the notion of the shadow as a semitransparent region in the image, which retains a representation of the underlying surface pattern, texture or colour.

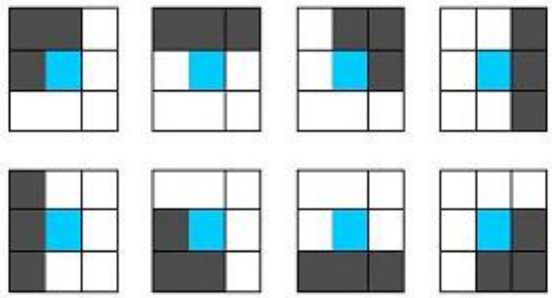
IV. COLOR CLUSTERING

From the result obtained from background subtraction we can see that the resulting contours of moving objects have been drawn roughly. But if we inspect the result carefully, it can be seen that there are at least two kinds of false segmentation lying near the contours of moving objects. The first is that background areas are falsely categorized to moving objects. The second is on the contrary. The reasons behind it may possibly be that the updating rate of background is not fast enough so the background model is not clean enough to extract the moving objects, and it may also be caused by image noise. These false segmentations will certainly degrade the accuracy of further processes, such as objects tracking, and be even worse when objects are close to each other. Moreover, some of these errors, which connecting with moving objects, can't be eliminated by general processing, such as smoothing, de-noising and erosion dilation based morphologic operations. In order to solve this problem without adding too much computational burden, we proposed a novel colour clustering based method as a post-processing to correct the false segmentations in the initial results Colour based image segmentation is a process of dividing an image into different regions such that each region has homogeneous colour. It is an important operation in many applications of image processing and computer vision, and has been extensively studied [18]. With colour based image segmentation, it can provide relatively complete boundaries of objects. Experiences tell us that in most cases, neighbouring pixels with similar colours should belong to the same objects, but the reverse deduction may not be true. So the goal of segmentation is to split each image into regions that are likely to belong to the same object. These regions or segments should be as precise as possible to distinguish the foreground objects from the background areas. There are many algorithms existing in the literature, for the sake of real-time characteristic, in this paper, we have tried two methods: K-mean algorithm implemented in OpenCV[9] and the method of over-segmentation[19,20]. The main difference between these two methods is the size of segment. The effect of using large segment is that it may straddle more than two objects or between the object and background area. It is undesirable. On the other hand, if the segment is too small, it may not provide sufficient information to distinguish the object from the other object or background. The use of over-segmentation strikes a good balance between providing segments that contain enough information for distinguishing and reducing the risk of a segment spanning multiple objects or over the background and the foreground area.

OVER-SEGMENTATION

The segmentation algorithm has two steps [19].

1. Image is smoothed using a variant of anisotropic diffusion. The purpose of smoothing is to remove image noise.
2. Then, the image is segmented based on neighboring color values.



SAD ALGORITHM

The smoothing algorithm iteratively averages along one of the eight directions as shown in Fig.2 The direction is determined by which direction has the minimum sum-of-absolute-differences (SAD) in color from the center pixel. After smoothing, each pixel is assigned its own segment. Two neighboring 4-connected segments are merged if the Euclidean distance between their average colors varies by less than a threshold (in [19], the value is 66). If the segment is too small, it will be merged with their most similarly colored neighbors. And if the segment is too large, it will also be divided. For more details, please refer to [19]. A result of the over-segmentation algorithm can be seen in Fig. For comparison, images with the averaged color value per segment of K-means and over-segmentation are also illustrated in Figure. Form the results comparisons in Figure 6, we can see that the result of over-segmentation looks more naturally. The reason is that the number of segments in the over-segmentation result is much larger than that of K-means. Though we can enforce the K-means algorithm to cluster more colors, but the price of time-consuming will increase greatly. On the contrary, the over-segmentation algorithm can run very fast, it is important for video surveillance applications.

K-MEANS

K-Means algorithm is an unsupervised clustering algorithm that classifies the input datapoints into multiple classes based on their inherent distance from each other. The algorithm assumes that the data features form a vector space and tries to find natural clustering in them. The points are clustered around centroid which are obtained by minimizing the objective where there are k clusters Si, i = 1; 2; : : : ; k and _i is the centroid or mean point of all the points

As a part of this project, an iterative version of the algorithm was implemented. The algorithm takes a 2 dimensional image as input. Various steps in the algorithm are as follows:

1. Compute the intensity distribution(also called the histogram) of the intensities.
2. Initialize the centroids with k random intensities.
3. Repeat the following steps until the cluster labels of the image does not change anymore.
4. Cluster the points based on distance of their intensities from the centroid intensities.
5. Compute the new centroid for each of the clusters.where k is a parameter of the algorithm (the number of clusters to be found), i iterates over the all the intensities, j iterates over all the centroids and are the centroid intensities.

I. EXPERIMENTAL RESULTS

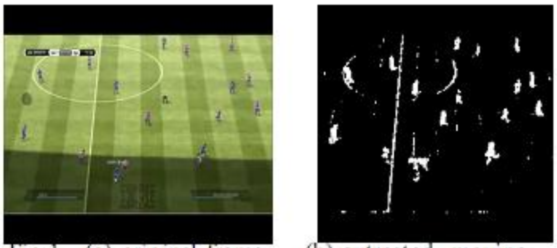


Fig.1 (a) original frame (b) extracted moving regions by mixture of Gaussians

Fig.2 (a) original frame (b) image after edge smoothing

Fig.3 (a) over-segmentation result (threshold 6) (b) over- segmentation result (threshold 66)

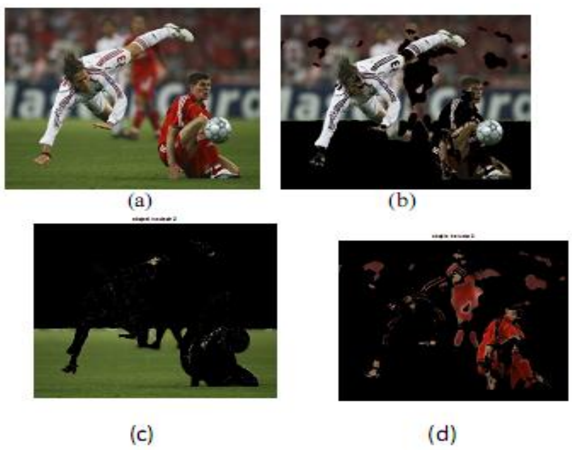


Fig.4 (a)original image (b) pixels in k-means cluster 1 (c) pixels in k-means cluster 2 (d) pixels in k-means cluster 3.

VI. CONCLUSION

Moving objects detection and segmentation is a fundamental step in many applications based on vision. Mixture of Gaussians is the frequently used method to subtracting moving objects from background. But its results are not good enough in some cases. In this paper, a post-processing method is proposed to solve this problem. The results with more complete boundaries provided by the color clustering is used to verify the outputs of mixture of Gaussians, and thus two possible false segmentations can be corrected effectively. Moving shadow suppression and small region filter are also adopted. Using these methods, the results can be greatly improved. Experiments have been done to prove the effectiveness of our work. As a general post-process procedure, the proposed method can also be used for other background subtraction related methods and the results can be used in next step-moving objects tracking.

REFERENCES

[1] C Niu, Y Liu, Moving object segmentation in the H. 264 compressed domain, Lecture Notes in Computer Science,vol 5995,2010:645-654.

[2] W Wang, J Yang, W Gao,Modeling Background and Segmenting Moving Objects from Compressed Video, IEEE Transactions on Circuits and Systems for video technology, Vol. 18, No. 5, 2008:670-681.

[3] Jens Klappstein, Tobi Vaudrey, Clemens Rabe, Andreas Wedel at el,Moving Object Segmentation Using Optical Flow and Depth Information,Lecture Notes in Computer Science, Vol 5414,2009:611-623

[4] Piccardi M. Background subtraction techniques: a review. IEEE International Conference on Systems, Man and Cybernetics, 2004, vol.4: 3099- 3104.

[5] Stauffer C, Grimson W.E.L. Adaptive background mixture models for real-time tracking. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Ft. Collins, 1999: 246-252.

[6] Horprasert T, Harwood D, Davis L S. A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection. Proceedings of IEEE ICCV' 99 Frame-Rate Workshop, 1999, pp.1-19.

[7] KaewTraKulPong P., Bowden R. An Improved Adaptive Background Mixture Model for Realtime Tracking with Shadow Detection. In Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems, Sept 2001, Pages:1-5.

[8] Elgammal A., Harwood D., Davis L. Non-parametric Model for Background Subtraction. in Proc. 6th Eur. Conf. Computer Vision, vol. 2, 2000, pp. 751-767.

[9] Dar-Shyang Lee. Effective Gaussian mixture learning for video background subtraction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(5):827 - 832.

[10] KaewTraKulPong P., Bowden R. An improved adaptive background mixture model for real-time tracking with shadow detection. in Proceedings of the 2nd European Workshop on Advanced Video-Based Surveillance Systems, Sept. 2001.

[11] http://www.cvg.rdg.ac.uk/PETS2009/a.html

[12] Intel Open Source Computer Vision Library. URL http://www.intel.com/ research/mrl/ research/opencv/.

[13] Gao X., Boult T., Coetzee F., Ramesh V. Error analysis of background adaption. in Proceedings IEEE conference on computer vision and pattern recognition, 2000, vol.1, pp. 503-510.

[14] Power P. W., Schoonees J. A. Understanding background mixture models for foreground segmentation. In Proceedings Image and Vision Computing, 2002, pp:267-271.

[15] Lee D.S., Hull J., Erol B. A Bayesian framework for gaussian mixture background modeling. in Proceedings of IEEE International Confererence on Image Processing, 2003, pages:973-976

[16] Mittal A., Huttenlocher D. Scene modeling for wide area surveillancd and image synthesis. in Proceedings IEEE conference on computer vision and pattern recognition, 2, pp. 160-167, (Hilton Head Isand, SC), June 2000.

[17] Cucchiara R., Grana C., Piccardi M., Prati A., Sirotti S. Improving shadow suppression in moving object detection with HSV color information. Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE, 25-29 Aug. 2001 Page(s):334 - 339.

[18] Lucchese L., Mitra S. K. Color image segmentation: A state-of-the-art survey. in Proc. Indian National Science Academy(INSA-A), vol. 67, A, New Delhi, India, Mar. 2001, pp. 207–221.

[19] C. Lawrence Zitnick, Sing Bing Kang, Matthew Uyttendaele, Simon Winder, Richard Szeliski. Highquality video view interpolation using a layered representation. Proceedings of ACM SIGGRAPH 2004, Pages: 600 – 608.

[20] C. Lawrence Zitnick, Sing Bing Kang. Stereo for Image-Based Rendering using Image Over-Segmentation.International Journal of Computer Vision, 2007, 75(1):49–65.

◈◈◈