Integration of fuzzy Markov random field and local information for separation of moving objects and shadows

Badri Narayan Subudhi, Susmita Ghosh, Sung-Bae Cho, Ashish Ghosh

Abstract

Major contribution of this article is to devise an efficient moving cast shadow segmentation technique that separates out the moving objects from their shadows casted on the background. It follows two major steps: background separation and shadow detection. For background separation, initially a background model is built. For a particular pixel location we construct the background model by taking the median of pixel values at the corresponding pixel locations in the temporal direction. To suppress the effects of quick change in illumination, and color frequency variation of texture background, in the proposed scheme we have extracted the RGB color features and ten local features at each pixel location in the considered target image frame and the constructed reference image frame. For background separation, a difference image is generated by considering pixel by pixel absolute difference of the thirteen dimensional target image frame and the constructed background model. This is followed by a spatial MRF constrained fuzzy clustering to find the moving regions in the considered scene. The maximum a posteriori probability (MAP) estimate of the fuzzy statistic based MRF are obtained by fuzzy clustering. The MAP of the MRF constrained fuzzy clustering provides a binary image, where the moving objects with the moving cast shadow are identified as one group and the background is obtained as another group. To segment the moving object from its shadow we explore a three stage shadow analysis technique. It uses analysis of rg chrominance property of shadow, local gray level co-occurrence based shadow processing followed by boundary refinement to separate out the regions corresponding to the moving cast shadow and moving objects. Performance of the proposed scheme is tested on several test video sequences. Effectiveness of the proposed scheme is verified by comparing the results obtained with those of some of the state-of-the-art techniques.

1. Introduction

Detection and tracking of moving objects are two important tasks in computer vision. It has numerous important applications: visual surveillance [35], face and gait-based human recognition [52], activity recognition [32], dynamic scene analysis [17], robotics [49], etc. It is observed that in vision systems, the capability of extracting moving objects from video is affected...
by noise, illumination variation, vegetation changes, shadow, etc. Since a shadow moves with the foreground object, static background segmentation techniques cannot differentiate them from moving objects. A shadow can create problem in moving object detection. Hence, detection and elimination of shadow from a video scene is an essential task.

A shadow in a scene appears when an object in the scene totally or partially occludes the direct light coming from the light source to reach the background. The part of an object that is not illuminated directly by light is called self-shadow, while the area on the scene blocked by the object (and not illuminated by light) is called cast shadow. The cast shadow has two parts: umbra and penumbra. The umbra corresponds to the area where the direct light is totally blocked; whereas, the area where light is partially blocked is called penumbra. Most of the shadow analysis techniques assume that umbra does not exist and it is a part of penumbra. If the object is moving, the cast shadow is known as moving cast shadow otherwise, it is called still shadow. In most of the cases, the moving cast shadow for a particular moving object in a scene occurs as a compact region, and is called compact shadow. It is quite common in a video processing system that the moving cast shadow will be detected as a moving object with a conventional background subtraction scheme [1,22,42].

A moving cast shadow detection scheme includes three fundamental steps: moving regions identification, shadow suppression and postprocessing. The former two are very important. The moving region identification stage uses different techniques: template matching [66], motion analysis [17], background subtraction (BGS) [3], etc. to identify the moving regions from different image frames of the considered video sequence. Among them the BGS technique is popularly used in shadow detection literature. Different BGS techniques like frame differencing [4], Gaussian mixture model [55], kernel based scheme [13], sparse local information based model [33], block alarm model [10], motion driven statistical model [30], fuzzy modeling [2,43], neuro-fuzzy model [5,6] etc., are popular and successfully used in the literature. The said scheme provides results which misclassify the moving cast shadow as a part of the moving objects. Hence, in the second stage of processing, a shadow detection technique is used to distinguish the shadow regions from the moving object regions. The post processing stage may finally detect compact shadow and moving object regions. The problem of shadow detection has attracted great interest in computer vision community.

The task of shadow detection becomes more critical if the shadow is cast on a textured surface [1]. The texture patterns are the results of physical surface properties such as roughness or oriented strands which often have a tactile quality, or they could be the result of reflectance differences (such as color) on a surface. In a real life video, textural patterns are very common and they also create problems in identifying the actual objects present in the scene. At the same time, the textural property of a region does not change if a shadow is cast on it.

In this paper, we propose an efficient moving cast shadow segmentation technique (for the sequences captured by fixed camera) that separates the shadows of the moving objects cast on a textured background. The algorithm follows two major steps: background separation and shadow detection. For background separation, initially a background model is built. For a particular pixel location, the background model has been constructed by taking the median of pixel values at the corresponding location in temporal direction. This gives a stable background model. For moving object detection we have used an absolute difference of the target image frame and the constructed reference background model followed by a spatial MRF constrained fuzzy clustering technique [8]. For spatial constrained fuzzy clustering, the RGB color and textural features (local) [20,53] are used. The MAP estimate of the fuzzy statistics based MRF are obtained by fuzzy clustering. The MAP of the MRF constrained fuzzy clustering scheme provides a binary image, where the moving objects with the moving cast shadows are identified as one class and the background as another one. To segment the moving objects from their shadows, a three stage shadow processing technique is followed. The technique uses an analysis of rg color chrominance property of shadow [64], local GLCM feature based shadow processing followed by boundary refinement to find the moving cast shadows and the moving objects in a scene.

The proposed scheme is successfully tested on different video sequences (indoor and outdoor) both for object detection and shadow removal, to prove its effectiveness. Different video sequences considered are with different background textures like grass field, road, floor carpets, wall and speckling water. It is observed that the proposed scheme effectively segments object in presence of visible and invisible shadow with compact shadow region. The performance of the proposed shadow detection technique is tested on several test video sequences. To validate the proposed scheme, results obtained by it are compared with those of the texture integration based shadow estimation technique [59], the adaptive shadow estimation technique [11] and gradient correlation based shadow estimation scheme [47,48]. The effectiveness of the proposed scheme is evaluated by three performance evaluation measures: shadow detection rate, shadow discrimination rate and combined score [11,42]. A preliminary experiments of the proposed work is reported in [57].

The organization of this paper is as follows. Section 3 presents the block diagram and the overview of the proposed technique in detail. The proposed background subtraction scheme used for separating the moving regions from the textured background is described in Section 4. The proposed shadow detection technique is narrated in Section 5. The data set used in the experiments and the results are analyzed in Section 6. Finally, Section 7 draws the conclusions and future work.

2. Literature on shadow detection

A recent survey on shadow detection techniques [48] has summarized the shadow detection taxonomy into four types: chromaticity based, physical property based, geometry based and texture based.

Chromaticity based shadow detection techniques use chrominance property of light or different invariant color models [14] to find out the characteristics of shadow from a given video scene. Salvador et al. [46] have studied a robust object and shadow separation scheme using color invariance property of shadow. The authors have used $c_1c_2c_3$ (a photometric invariant color features extracted from RGB color model) color invariant model in this regard. A robust moving object detection scheme that detects the
moving objects by eliminating the effects of shadow is proposed by Yang et al. [64], where, the authors have considered color, shading, and temporal consistency for shadow analysis. They have adopted normalized rgb color model for this. It is known that not a single color model is equally good for eliminating different kinds of shadows. Hence, it is a difficult task to determine the optimal color space in which the effect of shadow can be removed. Shan et al. [50] have made an extensive study on selection of optimal color space for a particular type of shadow elimination among a set of eight different color spaces. As per the spectral characteristics of shadow, the authors have divided a shadow into two categories: visible and invisible. If a pixel is in the visible shadow, the color value of that pixel will be much smaller than the actual background value. For invisible shadow, the pixel value is only smaller than the actual background scene. Their study reveals that HSV [12], $c_1c_2c_3$ [46], normalized rgb [15], and normalized rg [64] spaces are suitable for visible shadow detection, and YCbCr and $L=a+b^*$ [42] spaces are suitable for invisible shadow detection. However, it may be noted that shadow detection by chromaticity based techniques are affected by noise and can provide blob like structures in separating the shadow and the moving objects.

To overcome the disadvantages of chromaticity based shadow detection process, Nadimi and Bhanu [38] proposed a physical property based shadow detection technique that utilizes albedo ratio test and dichromatic reflection model. The approach was successfully tested on various image sequences with different surface roughness, and surface orientations. A semisupervised learning technique to segment the moving objects from their shadows is proposed by Joshi and Papanikolopoulos in [23], where the authors have used the characteristics of differences in color and edges in the video frames to detect the moving objects from shadow. To discriminate the moving pixels from the stationary shadow pixels, Mitra et al. [37] have proposed a robust shadow detection scheme, where three different threshold selection strategies, namely, $3\sigma$-rule, Hampel identifier, and adhoc threshold selection techniques have been used. Recently, Lalonde et al. [26] proposed a conditional random field based shadow detection technique using the concept of local cues of shadow. Here a classifier is trained to distinguish the ground shadow from noisy and blurred photographs. However, spurious edges on the shadow boundary and albedo test is not always reliable. In this regard, region based information obtained from image segmentation process is used for higher accuracy [19].

Geometry of a shadow can also be considered as a feature for shadow segmentation [48]. A lane dividing scheme is used by Hsieh et al. [21] to separate out moving objects and shadow from a given video sequence. This scheme uses both parallel and vertical line scanning systems as geometrical shadow feature to detect shadow from a given video sequence. Recently, a Gaussian mixture model based background subtraction scheme followed by edge analysis for moving vehicle and its shadow detection is reported by Lin et al. in [31]. The authors have shown the effectiveness of their algorithm on an application to intelligent transportation system.

From the above analysis one may conclude that shadow detection is an important task in computer vision. The problem becomes more critical if the shadow is cast on a textured surface. Different textural features were utilized in the literature for detecting shadow from the textured surfaces [48]. In a textured surface, it is very difficult to identify the actual boundary of the shadow due to high and quick changes of illumination, and color frequency variation [60]. In this regard, Tian et al. [59] have proposed a shadow detection technique using gradient information [28] based texture similarity measure. Sanin et al. [47] have proposed a shadow detection technique based on gradient correlation measure, where it is found that the shadow regions (corresponding to the background) in the target image frame should have a high texture correlation with the corresponding region in the reference image frame. An efficient use of Gabor filter is demonstrated by Leone and Distante for shadow detection [27]. They have used the similarity between the features from the textured patches obtained by Gabor filter in the background and the target frames. Recently, Choi et al. [11] have also proposed a three step shadow estimation scheme that uses chromaticity, brightness and local intensity ratio. The approach is tested on several textured sequences and is found to provide good results.

To the best of the authors’ knowledge, in the literature there is no method available that simultaneously takes advantage of the benefits of both gray level co-occurrence matrix (GLCM) [20] and fuzzy-Markov random field (MRF) [8] model for moving object detection and its shadow analysis. It also may be noted that the GLCM feature is also never used for shadow detection earlier.

3. Proposed scheme

A block diagram of the proposed technique is shown in Fig. 1. The first part of the block diagram represents the background construction block. The background model for a particular pixel location is constructed by taking the median of the pixel values.
of the said pixel location in temporal direction. For further processing, we have computed the local (Haralick’s) features at each pixel location of the target (from which the object is required to be detected) as well as the constructed background model/reference image frame. An absolute difference between the features of the reference image frame and those of the target image frame is calculated to obtain a difference image. By classifying this multi-spectral difference image into changed and unchanged classes [16], it is possible to detect the moving regions in the target image frame. To classify the difference image into two classes, we have used fuzzy Gibb’s Markov random field (GMRF) [8] to model the spatial attributes of the difference image. The binary segmentation problem is solved using the MAP estimation principle and the corresponding MAP is estimated in fuzzy clustering framework. It is found that the MAP of the fuzzy statistics based MRF provides a binary map and the objects with moving cast shadow are identified as one class and the background is obtained as another. To segment the moving objects and moving cast shadows into two different regions, we explore the merits of a three stage shadow processing technique. The steps are: analysis of chromaticity property of shadow, local GLCM feature based shadow processing and boundary refinement.

4. Background separation from moving regions

In this section, we describe the details on extraction of moving regions from the static background. In this regard, we have provided the details on background construction, feature extraction by GLCM, difference image generation and background separation by MRF constrained fuzzy clustering methods. A brief description on the basic theory of MRF for binary change detection is also provided in this section.

4.1. Background construction

In this approach, time instant and frame instant are assumed to be the same. Let the observed video sequence be a 3-D volume and $y_t$ represents the target image frame at time $t$. In the proposed scheme, we adhere to Lo and Velastín’s background modeling technique [34] for construction of the background model, as it gives a stable background [34]. We have characterized the background model for a particular pixel location by considering the median of the corresponding pixel location in the temporal direction. At a particular pixel location $(a, b)$, we have constructed the background model $B_{(t-1)}(a, b)$, considering the frames upto time $(t-1)$ as:

$$B_{(t-1)}(a, b) = \text{median}\{y_k(a, b), \ k = 0, 1, 2, \ldots, (t-1)\},$$

(1)

where, $B_{(t-1)}(a, b)$ represents the pixel value of the background model or reference image frame at location $(a, b)$ upto $(t-1)$th time instant.

4.2. Feature extraction and difference image generation

Selection of suitable features plays a critical role in object detection and tracking [66]. For analyzing on textured surfaces, extraction of textural features is important. Texture is a measure of the intensity variation of surface which quantifies properties such as smoothness and regularity. In the proposed scheme, we have considered the color features ($R, G, B$) along with ten local features [20,53]. Haralick’s features are calculated from the GLCM. A GLCM is obtained by constructing a 2D histogram which shows the co-occurrences of intensities in a specified direction and distance. The features are calculated from the average of the obtained GLCM matrices in four different directions namely, $0^\circ, 45^\circ, 90^\circ,$ and $135^\circ$. The distance parameter is selected to be greater than or equal to 1. In general, it is set to 1. In the proposed scheme, we have considered GLCM in a different way. We have computed GLCM in the target and the reference image frames considering a small window/mask over each pixel location. Hence, we term it as ‘local GLCM’. The size of the window/mask is assumed to be $n \times n (n = 7, 9, ...)$, depending on the size of the video image frames. A larger value of $n$ gives thicker object boundary with loss in shape of the actual object. While, smaller value of $n$ provides thinner object boundary with abrupt object edges. From the constructed local GLCM, different statistical features are calculated.

GLCM is a square matrix of dimension $N_g \times N_g$, where $N_g$ is the maximum number of possible gray levels present in the image. At a particular location $(i, j)$ of the GLCM, its value $p(i, j)$ is computed by counting the number of times a pixel with value $i$ has occurred adjacent to a pixel with value $j$ and the GLCM is then normalized. The ten different features considered in our work are: contrast, correlation, entropy, homogeneity, sum of square (variance), sum average, sum variance, difference variance and difference entropy [20,53]. For details on these features we refer to [7,20,24,53].

In our work, once the features are obtained, a difference image is generated by taking the absolute difference between the thirteen features (including $R, G, B$ and the ten local features [20,53] discussed earlier) computed at each pixel location in the reference and the target image frames. For a particular location $(a, b)$, the difference image $d_{i}$ is obtained as:

$$d_{i}(a, b) = |y_{i}(a, b) - B_{(t-1)}(a, b)|,$$

(2)

where $f$ is the number of features and it corresponds to $R, G, B$; and ten local features computed at location $(a, b)$. Here, the magnitude $d_{i}$ of the $f$-dimensional difference image is assumed to be a spatio-contextual entity. The process used for difference image generation in the proposed scheme is a multispectral difference image which is also a vectorial process of difference image representation. It may be noted that such difference image have independent spectral components and always reliably used in the literature for change detection [56].
4.3. Markov random field model for binary change detection

In this approach, the magnitude $d_i$ of the $f$-dimensional difference image is assumed to be a spatio-contextual entity. Each pixel in $d_i$ is assumed as a site $s$, $s \in S$, where, $S = M \times N$ represents the set of sites. $L$ represents a random field and $d_i$ is a realization of it. Let $W$ be a random variable and the realization of $W = \omega$ be a partition of the image into two region types, i.e., $\omega \in \{\omega_{ch}, \omega_{un}\}$ is a generic set of labels assigned to the pixels in $d_i$, where $ch$ corresponds to changed and $un$ to unchanged classes. The different classes present in an image can be spatially disjoint.

In MRF theory [29], it is known that the realization $\omega$ cannot be obtained deterministically from $d_i$. Hence, $\hat{\omega}$ is estimated from $d_i$. One way to estimate $\hat{\omega}$ is based on the MAP criterion, which yields a $\hat{\omega}$ value that maximizes the following posterior probability distribution

$$\hat{\omega} = \arg \max_{\omega} \{P(\xi = d_i|W = \omega, \theta)P(W = \omega, \theta)\}. \quad (3)$$

where $\theta$ is the parameter vector associated with $\omega$ and $P(W = \omega, \theta)$ is the prior probability of class labels. It is known that, if $W$ is an MRF, then it satisfies the Markovian property in the spatial domain [29]. According to the Hammersley–Clifford theorem [29], a given random field is an MRF if and only if its (joint) probability distribution $P(W = \omega, \theta)$ is Gibbs' distribution with $P(W = \omega, \theta) = \frac{1}{Z(\theta)} \exp \left\{-\frac{1}{\alpha} U(W)\right\}$, where $\alpha$ is the partition function expressed as $z = \sum_{\omega} e^{-U(W)}$ and $U(W)$ is the energy function expressed as $U(W, \theta) = \sum_{c \subseteq c} V_c(\omega)$, where $V_c(\omega)$ represents the clique potential function. $c$ represents a clique and $C$ represents the set of cliques. $T$ is the temperature constant and is considered to be $T = 1$ as in [29]. According to Pott's model, if the labels in a pair of neighboring sites are equal, then the clique potential function at site $s$ will be $V_c(\omega) = -\alpha$, else $V_c(\omega) = +\alpha$, where $\alpha$ is MRF model bonding parameter. Choice of parameter $\alpha$ controls the smoothness of segmentation. A larger value of $\alpha$ imposes a larger penalty for inhomogeneity. As $\alpha$ increases, the degree of smoothness imposed on segmentation also increases.

The likelihood function $P(\xi = d_i|W = \omega, \theta)$ models conditional dependency and is considered to follow Gaussian distribution. Accordingly, the parameter vector $\theta$ can be described as, $\theta = \{\Sigma_j, \mu_j, \pi_j\}$, where $\Sigma_j$ is the variance of the changed and the unchanged classes, respectively. Similarly, $\Sigma_{ch}$ and $\Sigma_{un}$ represent the covariance of the changed and the unchanged classes, respectively. $\pi_{ch}$ and $\pi_{un}$ are the point wise prior probabilities of the MRF for the site being in the changed and the unchanged classes, respectively. Thus, the likelihood function $P(\xi = d_i|W = \omega, \theta)$ can be expressed as

$$P(\xi = d_i|W = \omega, \theta) = \prod_{s \in S} \frac{1}{\sqrt{(2\pi)^f|\Sigma_j|}} e^{-\frac{1}{2} (d_i - \mu_j)^T \Sigma_j^{-1} (d_i - \mu_j)}.$$

Considering the prior probability of MRF and the likelihood function as in Eq. (4), the posterior probability of the MRF is written as

$$\hat{\omega} = \arg \max_{\omega} \sum_{s \in S} \left\{ A - \frac{1}{2} (d_i - \mu_j)^T \Sigma_j^{-1} (d_i - \mu_j) \right\} - \sum_{c \subseteq c} V_c(\omega). \quad (5)$$

where $A = -\frac{1}{2} \log ((2\pi)^f|\Sigma_j|)$ is a constant and $\hat{\omega}$ is the MAP estimate. However, such an approximation leads to a crisp estimation of the actual solution.

4.4. Fuzzy statistics based markov random field and MAP estimation

Gibb's Markov Random Field (GMRF) is a spatio-contextual statistical model mainly used for partitioning an image into a number of regions with the constraint of Gibb's distribution as a prior probability distribution [58]. In GMRF, the spatio property of the image can be modeled through different aspects, among which, the contextual constraint is a general and powerful one [29]. This is achieved by characterizing mutual influences among such entities in conditional probability framework. The important part of GMRF model is the heavy computational burden imposed by the model parameters estimation [8]. In the literature, Markov Chain Monte Carlo (MCMC) and Expectation Maximization (EM), pseudo-likelihood approaches are popularly used. Recently, cluster analysis methods are found to be popular. However, the cluster analysis method has main disadvantage of crisp decision for assigning a pixel into a particular cluster. One solution to such a problem is fuzzy clustering.

It is observed that MRF-MAP estimate using deterministic optimization gets affected as the scene is altered by real life vagueness and uncertainty. This uncertainty and vagueness may arise in the scene due to obscured edges, noisy environmental conditions, blurred scenes and low illuminated environmental conditions. Use of fuzzy sets [16,40] is found to provide satisfactory results in this regard. Utility of fuzzy sets for modeling object of interest in real life images is already established [8]. Use of fuzzy sets with local neighborhood and MRF model plays an important role in providing different amount of fuzziness for detection of the exact boundary of the moving object(s) in an image.

A modification of the posterior probability of MRF [29], incorporating the fuzzy statistics is done by Chatzis et al. [8]. Here, the authors considered a spatial penalty term in the objective function of fuzzy clustering so as to obtain a smooth transition in
the authors have made an equivalence between the objective function of the spatially constrained fuzzy clustering as fuzzy set as a probabilistic interpretation in MRF, they deduce the objective function as:

\[
\pi \text{ is the point-wise prior probability, mentioned earlier, and } K \text{ is the number of clusters. To obtain the spatial MRF constrained solution, the authors have made an equivalence between the objective function of the spatially constrained fuzzy clustering as expressed in Eq. (6) with the MAP of the MRF and hence, deduce a framework for MAP estimation. Imposing the concepts of fuzzy set as a probabilistic interpretation in MRF, they deduce the objective function as:}
\]

\[
J_m = \sum_{s=1}^{M \times N} \sum_{j=1}^{K} m_{sj} \text{dists}_{sj} + \lambda \sum_{s=1}^{M \times N} \sum_{j=1}^{K} m_{sj} \log \left( \frac{m_{sj}}{\pi_{sj}} \right).
\]

(6)

where, \( m_{sj} \) is the membership function for site \( s \) to the \( j \)th cluster. \( \text{dists}_{sj} \) represents the negative log-likelihood of the \( sth \) site belonging to the \( j \)th cluster and can be described as

\[
\text{dists}_{sj} = -\log(P(\mathbf{E} = di|W = \omega, \theta)).
\]

(7)

\( \pi_{sj} \) is the point-wise prior probability, mentioned earlier, and \( K \) is the number of clusters. To obtain the spatial MRF constrained solution, the authors have made an equivalence between the objective function of the spatially constrained fuzzy clustering as expressed in Eq. (6) with the MAP of the MRF and hence, deduce a framework for MAP estimation. Imposing the concepts of fuzzy set as a probabilistic interpretation in MRF, they deduce the objective function as:

\[
J_m = -\sum_{s=1}^{M \times N} \sum_{j=1}^{K} m_{sj} \log P(\mathbf{E}|W, \theta) + \lambda \sum_{s=1}^{M \times N} \sum_{j=1}^{K} m_{sj} \log \left( \frac{m_{sj}}{\pi_{sj}} \right).
\]

(8)

In our approach, the MRF based change detection is viewed as dividing the \( f \)-dimensional difference image \( (di) \) into two \( (K = 2) \) partitions. Allocation of pixels into different clusters is done by a hard decision (as in \( K \)-means clustering). To obtain the spatial MRF constrained binary segmented regions, we have considered the objective function as:

\[
J_m = -\sum_{s=1}^{M \times N} \sum_{j \in \{\omega_{ch}, \omega_{un}\}} m_{sj} \log P(\mathbf{E}|W, \theta) + \lambda \sum_{s=1}^{M \times N} \sum_{j \in \{\omega_{ch}, \omega_{un}\}} m_{sj} \log \left( \frac{m_{sj}}{\pi_{sj}} \right).
\]

(9)

At this point there is a need of estimating the parameters \( \mu_{ch}, \mu_{un} \) and \( \Sigma_{ch}, \Sigma_{un} \), where \( \mu_{ch} \) and \( \Sigma_{ch} \) correspond to the foreground object class and \( \mu_{un} \) and \( \Sigma_{un} \) correspond to the background class. Here, the EM algorithm with fuzzy clustering is used for parameter estimation.

In the present scheme, we have used the fuzzy clustering based MRF model for difference image modeling as in [8]. In this case we have also considered a multi-spectral feature based image rather than a simple color image and simple Markov model. Hence the proposed scheme uses MRF and FCM in a different way than it was used in [8].

5. Shadow segmentation

The MAP estimate of the fuzzy statistics based GMRF modeled difference image produces a binary image with two region types: foreground and background. It may be noted that the moving cast shadow also moves with the foreground objects and hence is identified as a part of the moving object. The said moving regions obtained by the MAP estimate (procedure described in previous section) as candidate shadow pixels, and it contains both moving objects and moving cast shadow pixels. A three stage shadow segmentation technique is proposed to identify or segment the moving cast shadow from the moving object. A block diagram of the proposed scheme is given in Fig. 2. The three steps of the scheme are analysis of chromaticity property of shadow, local GLCM feature based shadow processing, and boundary refinement. Each of these steps is discussed in the following sections. For detailed analysis on the proposed scheme we consider two examples and have provided the results of the proposed scheme for different steps of it. The considered video sequences are ‘Rahul’ and ‘Intelligent room’ video sequences. Fig. 3(a) represents one sample image frames from ‘Rahul’ and ‘Intelligent room’ video sequences. The region corresponding to the moving objects with moving cast shadows are shown in Fig. 3(b).

5.1. Analysis of chromaticity or color invariant property

The intensity of a pixel is an association of both illuminance and radiance component. The intensity value of a shadow pixel is different from that of the non-shadow. The chromaticity of the pixels in the shadowed area does not change much [36] and this is called color invariant property within a pixel [48]. In the proposed work, we assume that a significant change in pixel intensity value without any significant change in chromaticity is caused by shadow. The chromaticity of a pixel is obtained by transferring the pixel values in RGB color space to a well known normalized rg space. A pixel-wise comparison of the normalized...
Fig. 3. Steps of the proposed shadow segmentation scheme.

rg color information between the target image and the background image corresponding to the candidate shadow pixels helps in detecting the shadow pixels. In the proposed scheme, the following equations are used to transfer the color information in RGB to a well known normalized rg color space [64]:

\[
\begin{align*}
    r &= \ln \frac{R}{R + G + B}, \\
    g &= \ln \frac{G}{R + G + B},
\end{align*}
\]

where R, G, and B represent red, green and blue components, respectively. Similarly, r and g represent normalized red, and normalized green components, respectively. For each candidate shadow pixel at location \((a, b)\) we compared the normalized \(rg\) components of the background model and the target image frame using

\[
\Gamma(a, b) = \sqrt{\sum_{k=1}^{2} \left( y_{k,ca}(a, b) - B_{k(1),ca}(a, b) \right)^2},
\]

where \(k\) corresponds to two normalized \(rg\) color features obtained by Eq. (10). The subscript \(ca\) in Eq. (11) represents the pixel corresponding to the candidate shadow regions. A smaller value of \(\Gamma(a, b)\) (evaluated against \(k_1 \times \mu_\Gamma\)) signifies that the chromaticity of the pixel \((a, b)\) does not change much, and hence is more likely to be a shadow pixel. It can be represented as

\[
r_{(t+1)}(a, b) = \begin{cases} 
    1; & \text{if } \Gamma(a, b) \geq k_1 \times \mu_\Gamma \\
    0; & \text{otherwise},
\end{cases}
\]

where \(r_{(t+1)}(a, b)\) represents the output of the chromaticity based shadow pixels (0 corresponds to shadow pixels). The parameter \(\mu_\Gamma\) is obtained by taking average value of the \(M \times N\) sized \(\Gamma\) matrix and \(k_1\) as a constant which varies in the range [0.5, 1.5]. The region corresponding to the moving cast shadows obtained by the proposed chromaticity or color invariant property based shadow detection technique are shown in Fig. 3(c). It is observed from the chromaticity based shadow detection results that many parts of the moving cast shadow regions are identified as shadow. It is observed that some parts of the foreground object may be darker than the background. Thus, use of only chromaticity property for shadow detection is unreliable. Hence, the remaining part is not identified as shadow pixels. This part can be easily conceived from the results as shown in Fig. 3(c). Hence, the pixel \((a, b)\) for which \(r_{(t+1)}(a, b) = 1\) (also contains the moving object) is considered as the 1st set of candidate shadow pixels, and on these pixels local GLCM features will be used for further analysis.

5.2. Local GLCM feature based shadow processing

GLCM is one of the very popular tools for texture characterization. It is a matrix which shows the frequency of joint occurrence of gray values available in a defined neighborhood and is a spatio-contextual entity. The elements of the matrix are the function of angular relation and distance between the two neighboring pixels. The angles are quantized with 45° interval [53]. As is mentioned, the GLCM matrix is further used for calculation of Haralick’s features. These features are related to second-order statistics of image. In the proposed scheme, we have computed the GLCM at each pixel location by considering a mask/window around it as discussed in Section 4.2. At each pixel location of the GLCM, we have computed ten features. Among them, it is found out that the sum variance is a shadow invariant feature. Sum variance is a measure of dispersion of the gray level differences at a certain distance. Sum variance feature puts a relatively higher weight on the element which differs from the information content of both the row and the column of the GLCM. It may be noted that for the ambiguous texture window/region, in the GLCM we may obtain more entries of smaller magnitude as compared to a fewer entries of larger magnitude. Hence, sum variance feature will give a higher value for a homogenous region rather than a non-homogenous region and thus discriminates between the homogeneous and non-homogeneous regions. If a pixel from a particular scene undergoes shadow or illumination changes, its relative homogeneity does not vary much; on the contrary, a significant change occurs when the pixel is occluded by a moving object in the scene.

In the proposed scheme it is noticed that if a pixel intensity value is significantly altered due to shadow then the (local) sum variance texture feature computed at the said pixel location remains the same as compared to that belonging to non-shadow
region. Sum variance gives a higher value for homogenous regions rather than non-homogenous ones and it has the ability to discriminate between the two regions. Mathematically,

$$f_{svar} = \sum_{i=2}^{2N} (1 - f_{sent})^2 p_{row+col}(i).$$  

(13)

where $f_{sent}$ is the sum entropy and can be calculated as follows:

$$f_{sent} = -\sum_{i=2}^{2N} p_{row+col}(i) \log(p_{row+col}(i)).$$  

(14)

In the present work, we have compared the sum variance value calculated over each pixel location on the background model with that of the target image frame. If the resultant value is more than a predefined threshold, then the pixel is considered as a part of the moving object else it is considered to be casted by shadow. This operation is considered as:

$$\Delta(a, b) = |v^{ca}_{(svar,t)}(a, b) - b^{ca}_{(svar,t-1)}(a, b)|,$$

(15)

where the subscript $svar$ represents sum variance. The object and shadow are detected as

$$p_{(t+1)}(a, b) = \begin{cases} 
1; & \text{if } \Delta(a, b) \geq k_2 \times \mu_\Delta \\
0; & \text{otherwise},
\end{cases}$$

(16)

where $p_{(t+1)}(a, b)$ represents the output of the local GLCM feature based shadow processing. The parameter $\mu_\Delta$ is obtained by taking the average value of the $M \times N$ sized $\Delta$ matrix and $k_2$ is a constant in the range $[0.5, 1.5]$. The region corresponding to the moving cast shadow obtained by the proposed local GLCM based shadow processing scheme are shown in Fig. 3(d). It is observed from these results that most parts of the moving cast shadow regions and the moving object regions were identified. It is also observed that the foreground object detected by the GLCM based shadow processing sometimes provides output with irregular object and shadow boundary. It also can be observed from the results in Fig. 3(d). The pixels, which are identified as a part of the moving object by the local GLCM based shadow processing scheme, are labeled with 2nd set of candidate shadow pixels and is further treated with boundary refinement technique to obtain the final shadow pixels.

5.3. Boundary refinement

It rarely happens that in an illumination variation scene, the output of the GLCM based shadow processing provides two types of errors, namely, shadow detection failure and object detection failure. This may be due to the irregular object and shadow boundary obtained by GLCM shadow processing. The pixels which actually belong to shadow and are identified as object are termed as shadow detection failure. Similarly, the small regions which actually belong to object(s) and are identified as shadow are termed as object detection failure.

To improve the performance in this regard, we have followed a boundary refinement technique. If the shadow detected by an algorithm is an actual one, its boundary should be adjacent to the boundary of the obtained foreground candidates. It may be noted that accumulating this in a textured region using any heuristic is very difficult. It is also true that edge is a good feature to describe the texture. Hence, we have integrated the advantage of Sobel edge operation [18] in this regard. Sobel edge of the shadow candidate region (binary image obtained by fuzzy MRF-MAP framework and not the original image) is used to test whether the shadow region obtained by combination of chromaticity and local GLCM feature based shadow processing is a true shadow or not [54]. If the intersection of obtained shadow region and the candidate shadow (pixel) region boundary exceeds a predefined threshold value we consider the said region as a part of shadow. Thereafter, the number of boundary shadow pixels that are adjacent to the boundary of a foreground region ($N_s$) and the number of all the boundary shadow pixels ($N_c$) are found out. The shadow is considered as an actual shadow boundary if the ratio $N_s/N_c$ is more than 0.5. The shadow and the object region separated after the boundary refinement step are shown in Fig. 3(e) and (f).

In the proposed shadow detection process, the initial two stages i.e., analysis of chrominance property of shadow and local GLCM feature based shadow processing are very important stages. However we also have tested that alteration in order of their use does not affect the performance. In the proposed scheme we have used only $rg$ components of the normalized $rgb$ color space. It may be noted that if a shadow is casted on a very dark surface, the normalized blue components may give significant variation; and hence will provide an erroneous output. So, we have discarded the use of normalized $b$ (blue) components in our chromaticity analysis stage.

6. Results and discussion

Experimental results of the proposed algorithm are provided in this section. The ‘Results and discussion’ section is divided into three parts. In the first part we have provided a visual analysis of results obtained by the proposed scheme.

To evaluate the performance of the proposed algorithm, the results obtained by it are compared with those of the three different state-of-the-art techniques namely, texture integration based shadow estimation technique by Tian et al. [59], the adaptive shadow estimation technique by Choi et al. [11] and gradient correlation based shadow estimation scheme by Sanin
et al. [47,48]. We also compared the proposed technique with five other techniques. In the second part of the experiment, the performance of the proposed scheme is measured out by three performance measuring indices: shadow detection rate, shadow discrimination rate and combined score [11,42]. Ground-truth images are considered for evaluating the performance. A brief discussion on advantages and disadvantages of the proposed scheme is also given in this section.

6.1. Visual analysis of results

The algorithm is implemented in C and is run on Pentium D, 3.2 GHz PC with 4 GB RAM and Ubuntu operating system. The proposed scheme is tested on different video sequences, however for space constraint we have provided results on eight test video sequences. We have considered one aerial [9], six benchmark (one UCF50 action recognition) [44,61,62] and one real life video sequences. The six benchmark video sequences are Highway-I, Laboratory, Intelligent room, Campus, Walk with dog and CAVIAR.

In Carry video sequence, a person is moving over grass (texture). It is a low-resolution challenging aerial video sequence [9], where shadow of the person follows itself. The Intelligent room video sequence is an indoor sequence where a person is moving inside a hall. It is observed that the shadow of the moving person follows the person itself in the scene. Highway-I video sequence is an outdoor traffic scene, where the scene is affected by a moving shadow of the vehicles traveling on the road texture. The Laboratory video sequence, an indoor scene, is affected by the shadow of different moving persons in the lab. This is an example of light shadow. Rahul video sequence is a real life outdoor sequence, where the scene is affected by the shadow of a moving person ‘Rahul’. The Campus video sequence is a complex sequence with low illumination condition, where the scene is affected by the shadow of a moving vehicle and moving persons. Walk with dog is a complex sequence with real life grass texture where the scene is affected by the shadows of a moving person and his dog [44,61]. Similarly, the CAVIAR sequence is a real life hall monitoring sequence where the scene contains moving people and their shadows.

Fig. 4(a) presents a few sample frames of the considered sequences. All the results reported in this paper are randomly picked results of each dataset. Corresponding ground-truth images are provided in Fig. 4(b). The VOPs corresponding to the moving object and the moving cast shadow obtained by the texture integration based shadow estimation technique are shown in Fig. 4(c), where many false alarms are present. In Highway-I and Laboratory sequences, it is observed that the results have several missed alarms. The corresponding results obtained by the adaptive shadow estimation technique are depicted in Fig. 4(d). These results show that the moving object and the moving cast shadow in the scene are not properly segmented. Due to low resolution, major part of the moving object has been identified as shadow and hence, there is an increase in (shadow) false alarms. Due to textured background, many small object blobs are also misclassified in the scene. The results obtained by the gradient correlation based shadow estimation scheme are shown in Fig. 4(e) and are found to be better than those of the adaptive shadow estimation technique. However, many parts of the object are still misclassified in the scene and object & shadow are found to be over-segmented. Fig. 4(f) shows the moving object detection and moving cast shadow VOPs obtained by the proposed scheme. Comparing these results with those of other three mentioned techniques, it is seen that the moving object and moving cast shadow are detected in a better way using the proposed methodology.

6.2. Quantitative analysis of results

To provide a quantitative evaluation of the proposed scheme, we have provided three ground-truth based evaluation measures: shadow detection rate, shadow discrimination rate and combined score [11,42]. For all these measures we have initially built some manually segmented ground-truth images. The results obtained by the proposed scheme, the texture integration based shadow estimation technique [59], the adaptive shadow estimation technique [11] and the gradient correlation based shadow estimation scheme [47,48] are compared with the corresponding ground-truth images. For evaluating the accuracy, we have used the pixel by pixel comparison of the ground-truth images with the obtained results.

It may be noted that for better shadow detection, the measures shadow detection rate, shadow discrimination rate and combined score should be high. To validate the proposed scheme we have reported here the average of these measures obtained by taking the mean value over all the image frames of the considered video sequences. A comparative evaluation of the considered techniques for Carry, Intelligent room, Highway-I, Laboratory, Rahul, Campus, Walk with dog and CAVIAR video sequences are put in Table 1. We have reported the performance measures for all the considered video sequences with a fixed set of parameters ($\alpha = 0.8$, $k_1 = 0.70$, $k_2 = 0.72$) and also with the best set of tuned parameters (reported in Fig 4). From this table it is clear that the performance of the proposed method is superior to the other considered existing methods. For more rigorous evaluation of the proposed technique, we have provided the precision vs. recall curve for all the considered techniques in Fig. 8.

6.3. Analysis of background subtraction results with their evaluations

To validate the proposed background subtraction scheme, we have successfully tested it on several benchmark test video sequences including all the sequences of [41]. For visual illustration we have provided results for three sequences including two sequences from [41]. The considered sequences are Lobby [41], Underwater [63] and Water surface [41] video sequences. To test the effectiveness of the proposed BGS scheme, we have compared the results obtained by the proposed BGS scheme with those of the GMM based BGS [55], codebook based BGS [25], Bayesian modeling based BGS [51], multi-histogram clustering [65] and fuzzy background modeling [67] schemes.
Fig. 4. Moving cast shadow and object detection for different video sequence: (a) original image frames, moving cast shadow and object VOPs obtained using (b) ground-truth, (c) Tian et al. scheme, (d) Choi et al. scheme (e) Sanin et al. scheme, and (f) the proposed scheme.

A few sample image frames of the considered video sequences are provided in Fig. 5(a). The moving object obtained by the GMM based BGS technique are shown in Fig. 5(b). It is found that due to noise and illumination variation, several background pixels are falsely identified as moving object; and thereby resulting in object-background misclassification error. The results obtained by the codebook based BGS scheme are provided in Fig. 5(c), where several parts of the object are missed and thereby missed alarms are produced. Similarly, the moving object detection results obtained by the Bayesian modeling based BGS scheme are provided in Fig. 5(d). It is seen from Fig. 5(d) that an improved result for object detection is obtained than those of the GMM based BGS and codebook based BGS schemes. However, still it contains a few missed alarms. The results obtained by multi-histogram with MRF model and fuzzy background model are shown in Fig. 5(e) and (f), which shows that better results were obtained as compared to other techniques. The results obtained by the proposed background subtraction scheme are shown in Fig. 5(g), where an accurate shape of the moving object has been obtained. Hence, the results obtained by the proposed scheme is found to be better than those obtained using the other techniques.

To provide a quantitative evaluation of the proposed scheme, we have evaluated three ground-truth based evaluation measures: precession, recall and F-measure. A comparative evaluation of the considered techniques for Lobby, Under water, and Water surface video sequences are put in Table 2. Table 2 represents the average of precession, recall and F-measure computed over all the considered video sequences using different techniques. Table 2 confirms that the proposed BGS scheme provides better results as compared to other considered techniques.
Table 1
Average shadow detection and discrimination rate for different video sequence.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Carry</th>
<th>Intelligent</th>
<th>Highway-I</th>
<th>Laboratory</th>
<th>Rahul</th>
<th>Campus</th>
<th>Walk with</th>
<th>CAVIAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>η</td>
<td>ξ</td>
<td>Avg.</td>
<td>η</td>
<td>ξ</td>
<td>Avg.</td>
<td>η</td>
<td>ξ</td>
</tr>
<tr>
<td>Texture integration</td>
<td>60.35</td>
<td>81.15</td>
<td>70.70</td>
<td>79.98</td>
<td>89.10</td>
<td>84.54</td>
<td>72.31</td>
<td>89.20</td>
</tr>
<tr>
<td>Adaptive shadow estimation</td>
<td>53.20</td>
<td>82.05</td>
<td>67.62</td>
<td>95.01</td>
<td>91.39</td>
<td>93.20</td>
<td>84.98</td>
<td>88.97</td>
</tr>
<tr>
<td>Gradient correlation</td>
<td>66.32</td>
<td>78.01</td>
<td>72.16</td>
<td>90.35</td>
<td>97.32</td>
<td>93.50</td>
<td>80.50</td>
<td>94.11</td>
</tr>
<tr>
<td>Proposed (best)</td>
<td>84.03</td>
<td>91.12</td>
<td>87.57</td>
<td>96.11</td>
<td>98.07</td>
<td>97.09</td>
<td>87.32</td>
<td>95.20</td>
</tr>
<tr>
<td>Proposed (const.)</td>
<td>79.88</td>
<td>89.20</td>
<td>84.54</td>
<td>94.78</td>
<td>98.03</td>
<td>96.40</td>
<td>84.11</td>
<td>91.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B.N. Subudhi et al. / Information Sciences 331 (2016) 15–31
Fig. 5. Moving object detection for different video sequence: (a) original image frames, moving object detection using scheme: (b) GMM, (c) Codebook, (d) Bayesian modeling (e) Multi-histogram and MRF, (f) Fuzzy background and MRF and (g) proposed scheme.

Table 2
Average precision, recall and F-measure for different video sequence.

<table>
<thead>
<tr>
<th>Approaches</th>
<th>GMM</th>
<th>Codebook</th>
<th>Bayesian</th>
<th>Multi-histogram</th>
<th>Fuzzy background</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.64 ± 0.026</td>
<td>0.59 ± 0.011</td>
<td>0.79 ± 0.011</td>
<td>0.80 ± 0.044</td>
<td>0.83 ± 0.030</td>
<td>0.86 ± 0.021</td>
</tr>
<tr>
<td>Recall</td>
<td>0.82 ± 0.010</td>
<td>0.82 ± 0.070</td>
<td>0.81 ± 0.010</td>
<td>0.86 ± 0.060</td>
<td>0.87 ± 0.070</td>
<td>0.93 ± 0.030</td>
</tr>
<tr>
<td>F-measure</td>
<td>0.70 ± 0.010</td>
<td>0.69 ± 0.060</td>
<td>0.80 ± 0.010</td>
<td>0.83 ± 0.040</td>
<td>0.85 ± 0.080</td>
<td>0.89 ± 0.020</td>
</tr>
</tbody>
</table>

Fig. 6. Moving cast shadow and object detection for different video sequences of change detection Net database.

6.4. Results on other database

For better evaluation of the proposed scheme we have tested it on two standard databases: change detection net [1] and BMC [2].

The Change detection net is a benchmark database and widely used in different video processing applications for last few years. It contains several challenging sequences including shadow based sequences. In our work, we have used the different sequences under shadow categories. Similarly, the BMC is a benchmark BGS database, contains nine real life challenging sequences with different illumination condition.

The results of moving cast shadow and object detection results for all (six) sequences on shadow category of the change detection net database are shown in Fig. 6. It may be observed from visual results that the accuracy of shadow detection is quite good. For quantitative evaluation of the proposed scheme on this database, we have obtained the average shadow detection rate and the average shadow discrimination rate over all the video sequences and obtained the mean and variance values of each

measure. Quantitative analysis of these results reveals that the mean of the average shadow detection rate for all the six video sequences is 88.62% with 8.01% variance, the mean of the average shadow discrimination rate is 92.25% with 9.23% variance.

We also tested the proposed object detection scheme on BMC database. The proposed moving object detection results on few image frames of different video sequences (total nine) of BMC database are shown in Fig. 7. Results of the proposed object detection techniques were also evaluated by ground-truth based evaluation measures: precision, recall and F-measure. For all the sequences, we have obtained precision, recall and F-measure over all the video sequences and obtained mean and variance

Fig. 7. Moving object detection for different video sequences of BMC database.

Fig. 8. Precission vs Recall curves for different sequences (a) Carry, (b) Intelligent room, (c) Highway-L, (d) Laboratory, (e) Rahul, (f) Campus, (g) Walk with dog, and (h) CAVIAR.
values of each measure. Quantitative analysis of these results reveals that the mean of precision for all the nine video sequences is 83.11% with 7.55% variance, the mean of the recall is 90.12% with 8.11% variance.

6.5. Discussion and future work

Here the use of fuzzy concept incorporated MRF helped in providing accurate shape of the moving object from the scene where object in the scene has less variation from the background. Low illuminated environmental conditions, even can detect light shadows in the scene (as shown in Fig. 4). The proposed shadow detection technique uses GLCM based local features for shadow detection. The considered local features rely on sum variance (homogeneity property) of the local region, hence it gives better results against existing techniques like gradient/edge based correlation [47,48,59], contrast based shadow estimation [11], which enhances noise and blurred edges in the scene. It is also reflected from the experiments.

The parameters considered in the proposed scheme are: window size \((n \times n)\) for the extraction of local GLCM features, the chromaticity comparison parameter \(k_1\) and the sum variance comparison parameter \(k_2\). The window size \(n\) is chosen depending on the size of the video image frames. A larger value of \(n\) gives thicker object boundary with loss in shape of the actual object. A smaller value of \(n\) gives thinner object boundary with abrupt object edges. The parameter \(k_1\) varies in the range \([0.5, 1.5]\). A larger value of \(k_1\) misinterprets shadow as a part of moving object and smaller value of \(k_1\) misinterprets moving object as a part of shadow. Similarly, the parameter \(k_2\) varies in the range \([0.5, 1.5]\). A larger value of constant \(k_2\) misinterprets shadow as a part of moving object and smaller value of \(k_2\) misinterprets moving object as a part of shadow. Choice of parameter \(\alpha\) controls the smoothness of segmentation. A larger value of \(\alpha\) imposes a larger penalty for inhomogeneity. As the value of \(\alpha\) increases, the degree of smoothness imposed on the segmentation also increases. In the considered experiment the values of \(\alpha\) varies in the range \([0.1, 5.0]\). For better results, all the parameters are set according to some preliminary experiments (trial and error basis).

It may be noted that in the proposed scheme, we have considered the background model to be consisting of static background objects and have ignored the small vegetation changes like movements of leaves in a tree, water speckling, etc. As the proposed scheme relies on extraction of features from a local/small window or mask, the small vegetation changes will not affect the result so much. Let us consider an example of small vegetation changes due to the movement of leaves in a tree. It may be noted that tree is a static background object, but due to movement of its leaves few pixel values get altered. However this variation will cause a replacement of pixel values within its neighborhood only. Since the proposed scheme uses a small window for feature extraction, the movement of leaves will be within the neighborhood of the considered pixel and will not affect the result much.

For more rigorous evaluation of the proposed background subtraction and shadow detection technique and to have a fair comparison, we have tested them with two different combinations: proposed BGS with other shadow detection schemes, and GMM based BGS with the proposed shadow detection scheme. We found that a combination of GMM based BGS with proposed shadow detection scheme provides better results in detecting moving object and moving cast shadow as compared to the proposed BGS and other considered shadow detection techniques. However, the proposed BGS with other shadow detection schemes and GMM based BGS with the proposed shadow detection scheme do not provide better results compared to that of the proposed BGS and the proposed shadow detection technique together. We also compared the proposed technique with another combination i.e., the proposed BGS with local binary pattern (LBP) [39] based textural features for shadow detection in place of local GLCM based shadow detection. It is found that the proposed BGS with LBP based texture features some times fail to give good impression as there are two major drawbacks of it: a) LBP uses central pixel’s gray value as the threshold, hence it is sensitive to noise, b) a single change of binary level in the most significant bit may change the feature value to a great extend. We also have reported the average of shadow detection rate, shadow discrimination rate and combined score for all the video sequences with variances of these techniques for Curry, Intelligent room, Highway-I, Laboratory, Rahul, Campus, Walk with dog and CAVIAR video sequences are put in Table 3.

The average time consumed by various techniques are provided in Table 4. Here, we would like to mention that in the proposed scheme the GLCM based feature extraction takes a large computational time. It is observed from this table that among all the algorithms used in our study, texture integration based scheme takes the least amount of time for shadow detection. The proposed shadow detection scheme consumes comparable amount of time as in gradient correlation scheme and less time than the adaptive shadow estimation based one. This makes the proposed algorithm useful for specific offline applications, such as visual scene analysis, event analysis in surveillance, video annotation, video motion capture, pedestrian detection, shadow aware vehicle identifications, etc. All the results reported in this paper for different techniques were compiled in the local machine by the authors so as to give a fair comparisons of time.

<table>
<thead>
<tr>
<th>Approach</th>
<th>(\mu \pm \sigma_k^2)</th>
<th>(\mu \pm \sigma_k^2)</th>
<th>(\mu \pm \sigma_{k_hk})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prop. BGS+Tian et al.</td>
<td>74.06 ± 112.69</td>
<td>87.22 ± 28.28</td>
<td>80.64 ± 37.92</td>
</tr>
<tr>
<td>Prop. BGS+Choi et al.</td>
<td>77.22 ± 138.00</td>
<td>87.55 ± 15.76</td>
<td>82.37 ± 54.24</td>
</tr>
<tr>
<td>Prop. BGS+Sanin et al.</td>
<td>80.74 ± 40.93</td>
<td>89.37 ± 14.54</td>
<td>85.05 ± 18.37</td>
</tr>
<tr>
<td>GMM BGS+LBP feature</td>
<td>80.86 ± 77.85</td>
<td>87.95 ± 15.11</td>
<td>84.04 ± 34.41</td>
</tr>
<tr>
<td>GMM BGS+Prop. shadow</td>
<td>86.12 ± 31.60</td>
<td>89.27 ± 11.66</td>
<td>87.65 ± 14.50</td>
</tr>
<tr>
<td>Prop. BGS+Prop. shadow</td>
<td>89.04 ± 21.79</td>
<td>93.18 ± 10.36</td>
<td>91.16 ± 10.38</td>
</tr>
</tbody>
</table>

Table 3
Average shadow detection and discrimination rate for all video sequences.
In the proposed scheme, the boundary refinement process is used for proper refinement of the object-shadow boundaries only. In this process we take an intersection of boundary of the second set of candidate shadow pixel (which are identified as shadow by chromaticity and GLCM based shadow detection technique) and foreground pixel (obtained by the BGS technique). It may be noted that since the boundary refinement process uses the boundary of both the regions for shadow analysis, the edges of inner object/shadow pixels will not create disturbance in the final object detection. The boundary refinement process is considered for proper boundary refinement only.

In the proposed scheme we have considered all video scenes where background at a particular pixel location occurs for more than 50% times. It is difficult to estimate the actual background using the median based background estimation technique, from a video with cluttered environment, where background of the scene is occluded most of the time. Hence the proposed scheme is often found to be providing ghost in the scene. An improvement in object detection accuracy can be obtained by considering Reddy et al.’s [45] Markov random field based block wise spatio-contextual correlation based background estimation technique. The proposed technique is an unsupervised way of shadow detection. Here it may be noted that use of a classifier may be useful to do the same task. It needs some label patterns or ground-truth data to train the classifier. One way to provide the label information is to consider some confident label patterns from some thresholding process. However such a process is always not reliable to use and may lead to erroneous output.

It is observed that for large vegetation and sudden light changes the proposed scheme does not give satisfactory results. Hence in future we would like to use the local contrast with probability distribution function to handle such situations. It is also observed that the proposed scheme does not give good results for monochrome and moving camera captured video sequences. We will also like to work on these issues in future.

7. Conclusion

In this paper, an integration of fuzzy MRF and local GLCM based Haralick’s features were explored to segment/separate a moving object from its shadow cast on a textured background. In the initial phase, the moving regions in a target image frame is identified by taking an absolute difference of the 13 dimensional (RGB and ten local features) target image frames and the constructed background model followed by a fuzzy statistics based MRF clustering. This process detects the moving cast shadow and the moving object as moving regions. To separate the moving objects from their shadows an analysis of rg color chrominance property of shadow, local GLCM feature based shadow processing followed by boundary refinement is done. Performance of the proposed shadow detection technique is tested over several test video sequences. From the experimental results, we can conclude that the proposed scheme can be successfully applied for detecting moving objects by segregating the effects of their shadows. The proposed scheme is validated against the results obtained by those of the texture integration based shadow estimation technique [59], the adaptive shadow estimation scheme [11] and gradient correlation based shadow detection scheme [47,48], and is found to be better.

Acknowledgments

The authors wish to thank the anonymous referees for their constructive criticism and valuable suggestions for improving the quality of the manuscript.

References