Real Time Object Detection System Based on Color and Spatial Information

[1]Md. Zahangir Alom, [1]Syed Shakib Sarwar, [1]Rubel Biswas, [1]Moin Mostakim, [2]Hyo Jong Lee
[1]School of Engineering and Computer Science, BRAC University, Dhaka, Bangladesh.
[2]Dept. of Computer Science and Engineering, CAIIT, Chonbuk National University, South Korea
zahan.cse@gmail.com, shakib87@gmail.com, rubelbiswas@bracu.ac.bd, shisshir11@gmail.com, hlee@jbnu.ac.kr

Abstract— **This paper proposes a statistical color model of background extraction based on Hue-Saturation-Value (HSV) color space instead of RGB which shows better use of the color information. HSV color space corresponds closely to the human perception of color and it has revealed more accuracy to distinguish shadows. The key feature of this segmentation method is processing hue component of HSV on image. The components of HSV are efficiently analyzed and treated separately so that the proposed algorithm can adapt to different environmental illumination conditions and shadows. Polar and linear statistical operations have been applied to extract the background from the video frames. Morphological operations are used to reduce the noise of foreground of the input frame. Label matrix has been determined for finding the object region in the foreground image. The proposed system considered age (duration of object existence) of the objects, misses, potential detections and potential for meeting stationary objects criterions. Finally, detected objects have been indicated with yellow boundary and frame number. The experimental results show that the proposed foreground segmentation method can automatically segment video objects robustly and accurately in various illuminating and shadow environments.**

Keywords- tracking objects, polar statistics, label matrix, optimal threshold, Gaussian Model, foreground extraction.

I. INTRODUCTION

Background subtraction is widely used method for identifying moving and stationary objects from video stream. It is the first significant step in many computer vision applications, including video surveillance, human motion, monitoring traffic and analysis of suspicious event occurrences. The performance of these applications is dependent on the background subtraction algorithm being robust to illumination changes, small movements of background elements (e.g. swaying trees, under water imaging), the addition or removal of items in the background (e.g. parked car), and shadow cast by moving objects. Computational efficiency is also high priority as these applications generally aim to run in real-time. The most common paradigm for performing background subtraction is to build an explicit model of the background. Moving objects are then detected by taking the difference between the current frame and this background model. Typically, a binary segmentation mask is then constructed by classifying any pixel from a moving object when the absolute difference is above a threshold. Background subtraction algorithms differ in how they define and update the background model. Despite the success enjoyed by background subtraction algorithms, it is becoming clear that post-processing is required in order to improve their performance. This post-processing can range from shadow detection algorithms operating at the pixel level to connected component labeling which identifies object level elements. The results of post-processing can be used to directly improve the quality of the segmentation mask and feedback into the background subtraction algorithms in order to facilitate more intelligent updating of the background model.

This paper is organized as follows. Section II represents the related works on this field. The proposed pipeline algorithm is discussed in Section III. In this section an important concept of polar statistics and accurate process direction of hue data components is also presented. Section IV describes how color distribution can be approximated by parametric descriptions and how the adaptive background model distinguishes between foreground and background. Optimal threshold calculating procedure represented in this Section. Post processing is described in Section V and section VI describes on foreground objects and special information extraction procedure and experimental result and conclusion of this work are given in Section VII and Section VIII.

II. RELATED WORKS

There are seven methodologies which have been explored on background subtraction and comparative studies on those methodologies are discussed in this paper. This original review allows the readers to compare the methods' complexity in terms of speed, memory requirements and accuracy and can effectively guide them to select the best method for a specific application in a principled way [1]. Previous papers proposed automatic segmentation that is change-detection. It uses the luminance, color, texture or shape changes between frames to detect and segment the video objects. Change-detection can be further divided into adjacent frame subtraction and background subtraction. Though adjacent frame subtraction can adapt to luminance change with ease, it is hard to acquire the whole body of the video object due to the various movement of the video object. Previously most of the systems consisted with background subtraction technique for detection of moving objects on a scene of the video. Comparing of observed image with background reveals differences that are related with the objects of interest. However the biggest problem is how to correctly model the background image, especially when it is highly dynamic, like an aquatic environment. To overcome this problem some paper represented a simple technique that is to average the incoming images over time. Although, those method has problems in

detecting slow moving objects [1][9].In references [2]~[8], there are many color space based method proposed with respect to different experimental environments. Especially [3] presented comparative study on color base system and they used RGB, HSV, and YCrCb and normalized RGB. In this paper[3] they conciliated that YCrCb color space is suitable for the detection of foreground and shadow in traffic image sequences and next step of the proposed study will be to study how different background modeling techniques work with different color space. In references [4] ~ [8] used HSV color space and proved HSV color space can adapt to different environments, various illumination conditions and robust to shadow.

Ming Zhao et al [3] presented robust background subtraction system based on HSV color space. Our proposed system has some similarity with that proposed algorithm. But they did not consider any object detection concepts from finally segmented frame. An adaptive threshold calculation technique has been applied for extracting the foreground from the current frame, whereas, the previously published paper used a hypothetical approach. Moreover, Adrea Prati et al.[4] suggested inclusion of directions to include spatial information and post processing into the Sakbot system or to try ATON in the HSV color space. The Pfinder in reference [6] system in MIT used a method based on YUV color space. It performed well only with little gradual illumination changes. If the luminance changes enough then the results show very poor performance. Reference [7] presented a pedestrian tracking based on HSV color space and spatial information. François et al. [8] presented an HSV color space based background subtraction technique. It can produce good results. But it subtracted only the background frame from the current frame and results contained a lot of noise. And it did not analyze the different properties of each color component of the pixels and how to process them separately, which led to little robustness. Reference [9] represented a system for highly dynamic aquatic environments but they did not distinguish between people and objects in foreground frame after detecting objects. A different background segmentation techniques have been provided in [11] ~ [14] based on adaptive mixture of k Gaussians. However, this approach shows some problems related with the number of k classes to choose and which class of k should represent the background.

Also, Gaussian parameter estimation based on Expectation-Maximization (EM) algorithm is computationally extremely expensive. Object detection and tracking system is represented in [15] [16] in details. References [17][18] represents each pixel by the minimum, maximum and largest inter-frame absolute difference initially estimated from the first few seconds of video and are periodically updated for those parts of the scene which did not contain foreground objects. In [19] modeled foreground pixels using the mean and covariance, which are recursively updated. The methodology proposed is based on multivariate Gaussian mixture model for background, whose parameters are estimated through standard EM algorithm, on Hue Saturation Value (HSV) color space [20]. We have taken region based labeling concepts from [21]. However, comparing with the existing techniques, the proposed method has the following advantages: (1) this proposed system is able to distinguish between peoples and objects. (2) Able to detect the objects characteristics like moving or stationary. (3) Less noise and higher accuracy are achieved with shadow detection technique and by applying post processing algorithm. (4) Robustness is guaranteed by analyzing the different properties of each pixel's color components and their statistical features. (5) Again proposed system is able to represent the people or other objects with a rectangular box successfully. It will be shown, by qualitative and quantitative results, that our adaptive model, extending, can cope with all the above mentioned issues for foreground maintenance and achieves robust detection with boundary box for different types of videos taken with stationary camera [5][9].

III. PIPELINE OF ALGORITHM

A. Flow of processing: The block diagram of the proposed pipeline algorithm is illustrated in Figure 1. The proposed system consists with two main parts: one is background extraction and another is object detection. First four steps are related to background extraction and remaining steps are related to object detection. The frames have been converted from RGB to HSV color space. As we know the HSV color space consists with polar co-ordinate system for hue, so polar statistics is used to calculate the mean and variance of the hue components. Statistics on polarization is discussed in the subsection B.

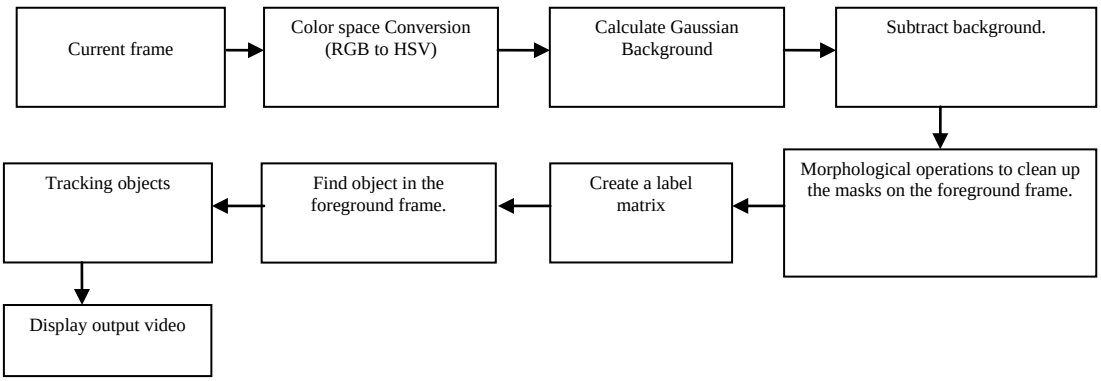


Figure1. Diagram of proposed pipeline algorithm.

B. Applied polar statistics: The algebraic structure of the line and the circle are different and therefore adequate methods of circular data analysis are discussed in [23]. It has been used for working with directional data. In contrast, to the linear domain only one operation is possible, the addition modulo 2π is available in the circular domain. Due to the fact that the circle is a closed curve, its natural periodicity must be taken into account. As described in [22] a set of N angular estimates can be represented by N unit phases with arguments equal to the corresponding angular estimates. The mean angle $\hat{\mu}_p$ is then given by the argument of the phasor sum and this value is independent of the choice of origin. The general definitions of the circular mean and variance based on this phasor sum are of the following from.

Definition: Circular sample mean and sample variance: Let $\{\hat{a}(k)\}, \hat{a}: Z \mapsto R$ be a set on N observation of a random variable in the circular domain $[0, P)$. Then the circular sample mean $\hat{\mu}_p$ and circular sample variance $\hat{v}_p$ are defined by

$$\hat{\mu}_p = \frac{P}{2\pi}\left(\arg\left[\sum_{k=0}^{N-1} e^{\frac{j2\pi a(k)}{P}}\right]\right)_{2\pi} \quad (1)$$

and $$\hat{v}_p = \frac{P^2}{4\pi^2}\left[1-\frac{1}{N}\left|\sum_{k=0}^{N-1} e^{\frac{j2\pi a(k)}{P}}\right|\right] \quad (2)$$

where $((\ ))_{2\pi}$ denotes reduction modulo 2π onto $[0,2\pi)$. The circular variance $\hat{v}_p$, $\hat{v}_p \in \left[0, \frac{P^2}{4\pi^2}\right]$ cannot be compared directly with its linear equivalent σ^2 which lies in the domain $[0,\infty)$. However by using the relationship between the normal distribution on the circle (wrapped normal, [9]) and the normal distribution on the line a circular standard deviation in the range $[0,\infty)$ can be define as

$$\hat{\sigma}_p = \sqrt{-2\log_n\left(1-\frac{4\pi^2}{P^2}\hat{v}_p\right)} \quad (3)$$

Therefore when using statistical definition in the context of hue values, we always refer to the above definitions from polar statistics. The calculation procedure of polar mean for hue of the proposed system is given below:

i) Take the pixel values for hue of $f_{hsv}(x_i, y_i)$, i=1……N and calculate the $\theta = \left(f_{hsv}(x_i, y_i)\right)\%(2\pi)$ (4)
where $x = \cos(\theta)$, and $x = \sin(\theta)$

ii) μ_x ,and μ_y are calculated from $f_{hsv}(x_i, y_i)$ respectively.

iii) If data is without dimension then

$$\hat{\mu}_p = \left(\tan^{-1}\left(\mu_y, \mu_x\right)\right)\%(2*\pi) \quad (5)$$

Else $$\hat{\mu}_p = \left(\tan^{-1}\left((\mu_y, d), (\mu_x, d)\right)\right)\%(2*\pi) \quad (6)$$

here d refers data dimension.

IV. The Method

A. Foreground extraction: The first step of developing a foreground $f_f(x, y)$ extraction is to build a model of the background $f_b(x, y)$. Since there are no preset background images to use, the software will have to generate a model automatically. Using the statistical approach, the software will build a Gaussian model. A Gaussian Model calculates each pixel-value from all the sample pixel's mean and variance. The model sets a lower bound and an upper bound that will eliminate pixels that are outside of the norm. If a video is to run for an extended period of time, the pixels average will equal to the background's values unless the foreground object stays static. Steps of the foreground extraction are as follows:

1. To Generate Gaussian background model in $f_{hsv}(x, y)$ color space for each pixel, smaller data range [0.45 0.55] is used to ensure faster calculation. Firstly Gaussian background mean (μ_b) and standard deviation (σ_b) for specific portion of data is calculated by using HSV Gaussian Model. We have Calculated of μ_{sv} from μ_s , μ_v and σ_{sv} from σ_s , σ_v for saturation and value of pixels' of the respective frame. The polar mean $\hat{\mu}_p$ for hue color pixels of the respective frame is calculated according to the Eq.5 and Eq.6 in section III. Now take frame $f_{hsv}(x_i, y_i)$, where i=1……N, to generate the new frame $h(x, y)$ only for hue pixel, we can express as follows
$$h(x, y) = \left(f_{hsv}(x, y) * 2 * \pi - \hat{\mu}_p + \pi\right)\%(2\pi) \quad (7)$$
$h(x, y)$ is used to calculate Gaussian mean (μ) and variance (σ) with respect to above range then
$$\mu_h = \left(\left(\mu - \pi + \hat{\mu}_p\right)\%(2*\pi)\right)/(2*\pi) \quad (8)$$
and $$\sigma_h = \sigma/(2*\pi) \quad (9)$$
The mean μ_b and deviation σ_b are calculated from μ_h , μ_{sv} and σ_h , σ_{sv} respectively. And Calculated means and deviation represent the background mean (μ_b) and background deviation (σ_b) respectively. Find the scaled deviation *(D)* of the current frame from background.
$$D = \sum_{i=1}^{N}\left(\left\| f_{hsv}(x, y)\right\| - \mu_b\right) \quad (10)$$
2. Compare with optimal threshold (T) to generate label, like $D>T$ for each deviation. Calculating procedure of optimal threshold T is discussed in the next section.
3. Apply the morphological operations like closing $(\bullet)$, opening $(\circ)$ and filling.
4. Find the largest connected component.
5. Considered the label as foreground of the frame that

is $f_f(x, y)$.

B. Determining the optimal threshold based on Gaussian model:

Thresholding technique is used in this step. Pixels deviation values that are higher than the threshold value considered as foreground and others as background. Setting the threshold value is very important for any system to detect foreground and background. For selecting this threshold we used optimal thresholding technique. The frame of the video contains some principle region and the distribution of pixel values in each region follows a Gaussian distribution. The proposed system uses the following algorithm for selecting the optimal threshold for this system as:

1. Find the histogram $h(z)$ of the normalized frame to be segmented.
2. Calculate the probability of a pixel value by the following mixture.
$$P(z) = P_b p_b(z) + P_f p_f(z) \quad (11)$$
where $p_b(z)$, and $P_f(z)$ are probability distributions of background and foreground pixels. And P_b, and P_f are priori or a posteriori probabilities of background and foreground pixels.
3. Overall probability of error:
$$E(T) = P_b E_f(T) + P_f E_b(T) \quad (12)$$
where $E_f(T) = \int_{-\infty}^{T} p_f(z)dz$ and $E_b(T) = \int_{T}^{\infty} p_b(z)dz$
4. Minimize $E(T)\dfrac{dE(T)}{dT} = 0$
5. The above expression is minimized when
$$T = \frac{\mu_b + \mu_f}{2} + \frac{\sigma^2}{\mu_b - \mu_f}\ln(P_f / P_b)(\sigma_b = \sigma_f = \sigma) \quad (13)$$
6. Special cases when $P_b = P_f$ or $\sigma = 0$,
$$T = \frac{\mu_b + \mu_f}{2} \quad (14)$$

T is the final optimal threshold of the propose system.

V. Post Processing

As the segmentation noise is inevitable, post processing is needed to refine the segmentation results. This phase performs filtering techniques to eliminate remaining noises. Noise can be small or big. Therefore, block labeling algorithm has been applied for refilling the objects region in the foreground frame. It can also fill the holes inside the segmented foreground frame. An algorithm proposed in literature [24] has been used to efficiently find all the blocks in the binary frames. For big blocks which are regarded as video objects, inside region have been filled. Finally morphological operations and edge rounding techniques are used to remove the noise and smoothing the edges respectively [9].

VI. Obeject Finding and Tracking

After subtracting the background from the video input frame successfully, object extracting techniques have been applied on foreground frame. The discussions on different region extracting techniques are given below.

A. Finding object region

Number of objects in the video frame, the area of the objects, centroid of the objects that the points in a system of message each of whose co-ordinate is a weighted mean of co-ordinate of the same dimension of points within the system, the weight being determined by the density function of the system and boundary region of the objects [15][16]. The following steps are used for finding the objects:

1. The area is defined to be the number of pixels belonging to the region.
2. Centroid calculation that is the mean of the pixel list of the region.
3. For calculating the boundary of the region, height (h) and weight (w) are calculated.

B. Object Tracking

Some parameters are used for object tracking from the foreground frame like area change fraction is the percent size that an object can vary, centroid change fraction is percent ratio to size that an object can move, maximum consecutive miss is the maximum number of occlude frames minimum persistence ratio is the ratio of persistence for relevance and alarm count is the minimum number of frames before an object can be considered abandoned. Algorithm as follows:

1. For each object in the current frame, scan through existing list to find match. If no match exists then add to our tracking list or object list.
2. If the object is already being tracked by looking through the existing object list for close match according to following process:

i) Calculate difference in area and centroid

ii) If the difference in area and centroid are small enough, then this object is already tracked. And update the information about this object being tracked.

iii) Else add to the object list in first available position.

iv) Fill element information about the new object.

3. End.

Process the tracking list or object list for age, misses, potential detections, and potential for meeting stationary objects criteria.

1. Only process the active entries in the objects list.
2. Keep track of object “misses” in the miscount field and clear the consecutive misses.
3. Mark the object for detection from the object list if

i) Consecutive misses exceeds maximum consecutive miss

ii) Ratio of hit count to age drops below minimum persistence ration

4. Output the bounding box of objects that have been stationary for more frames than “alarm count”.

VII. EXPERIMENTAL RESULTS

For evaluating the proposed system, video sequence with a 360x240 pixels resolution is considered. The experimental video frame rate and data rate was 30f/s and 65029kpbs respectively. Outdoor video at train station has been considered as experimental video. The experimental video was very complex because the train was running in the backward direction with respect to the peoples. Shadows have been created by the train as well others object.

The experimental video contains two types of shadow like self shadow and cast shadow. The proposed system successfully extracts the peoples from that complex background environment. Some false segmentation results have been shown on ground section. This false segmentation is caused by shadows. In the future, shadow removal techniques will be applied for removing the false segmentation in the pre-processing section. Fig.3 represented the segmentation result before and after applying morphological image processing operations. Figure 4 represented output objects with specific object boundary box.

The proposed system provides the rectangular box over the specific object it may be human or any other moving or stationary objects. It is clear from the experimental results that the proposed algorithm is still robust and can produce accurate results in bad illumination condition.

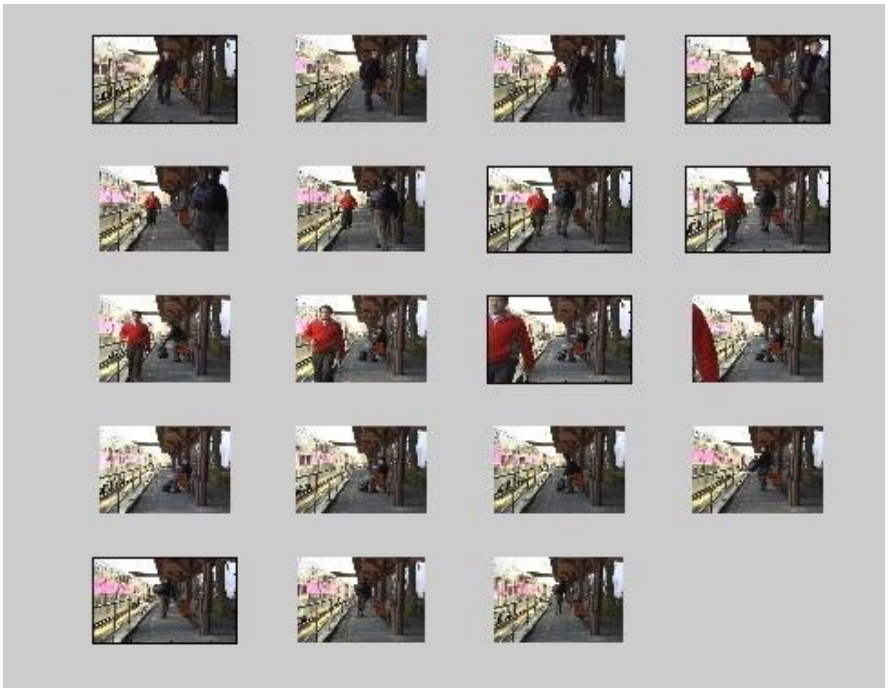
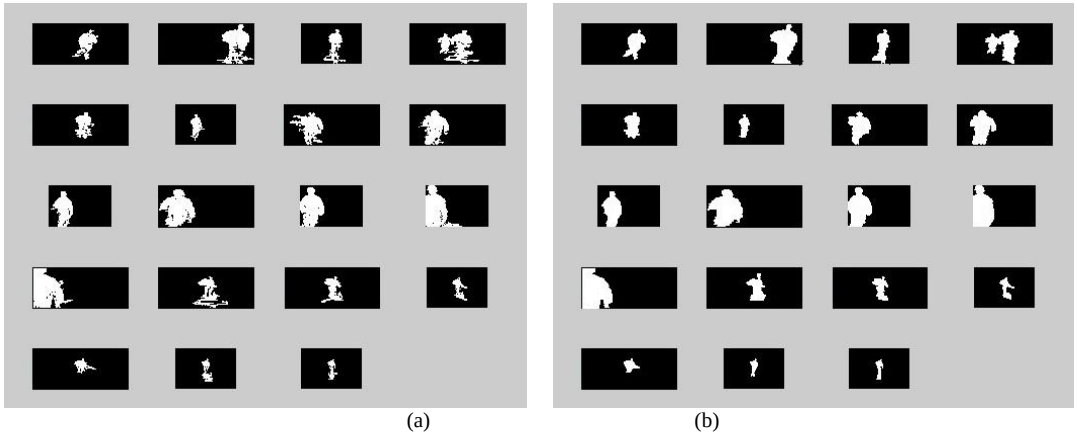


Figure 2. Input frames.



(a) (b)

Figure 3. (a) Primary foreground extraction results and (b) finally, segmented frames after applying morphological operations.

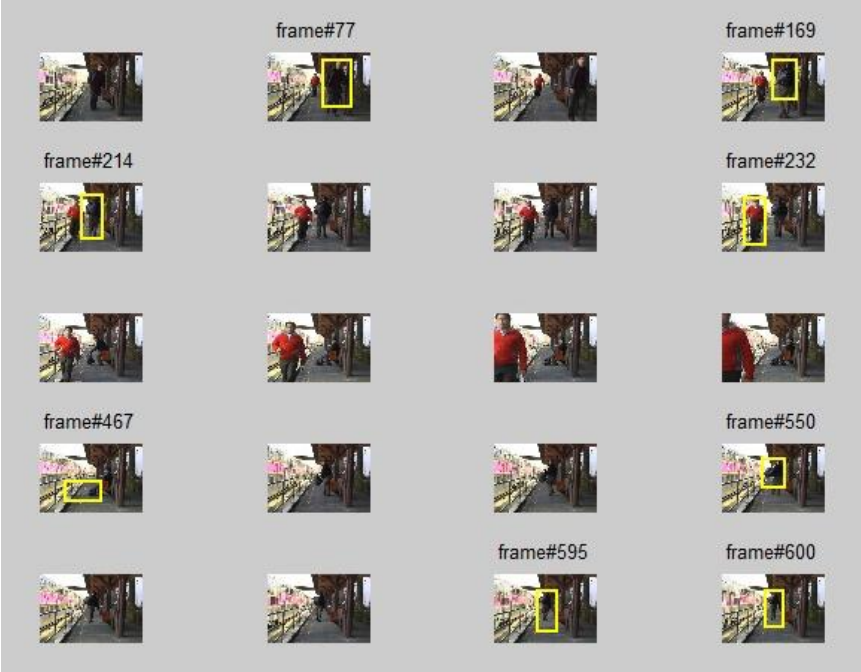


Figure 4. Object finding and tracking output. The proposed system provides the rectangular box over the specific object it

VII. Conclusion and Future Works

This paper presented the circular nature of hue in the HSV color space and provides accurate density functions for modeling color distributions for extracting foregrounds object from video frames. The statistical foreground extraction model and objects detection procedure based on calculating Gaussian means and variances of the pixels of the frames have been described in details. The experimental results represents that the proposed system can detect foreground objects successfully. In addition to that, this system also proposed an effective solution to recognize stationary or moving objects in the video frame. Shadow detection and post processing have been applied to refine the results. All the approaches contribute to the robustness of the foreground extraction method and experimental results are very satisfactory.

References

[1] Massimo Piccardi "Background subtraction techniques: a review", IEEE International Conference on System, Man and Cybernetics 2004.

[2] N. Herodotou, K.N. Plataniotis, and A.N. Venetsanopoulos, " A color Segmentation scheme for object-based video coding" in proceedings of the IEEE Symposium on Advance in Digital Filering and Signal Processing, 1998, pp.25-29.

[3] Pankaj Kumar, Kuntal Segupta, Adrian Lee "A Comparative Study of Different Color Spaces Foreground and Shadow Detection for Traffic Monitoring System", IEEE, 5th International Conference Transportation Systems 3-6 September 2002, Singapure.

[4] Adnrea Prati, Ivana Mikic, Constantino Grana and Mohan M. Trivedi "Shadow Detection Algorithm for Traffic Flow Analysis: a Comparative Study", IEEE Intelligent Transportation Systems Conference Proceedings- Oakland (CA), USA-August 25-29, 2001.

[5] Ming Zaho, Jiajun Bu, and Chun Chen "Robust background subtraction in HSV color space" in SPIE: Multimedia System and Applications V, Boston, USA, July 2002, Vol. 4861, pp. 325-332.

[6] C. Wren, etc. "PFinder: Real-Time Tracking of the Human Body", IEEE Transaction on Pattern Analysis and Machine Intelligence, pp.780-785, 1997.

[7] Florian H. Seitner and Brian C. Lovell "Pedestrian tracking based on color and spatial information" IEEE Proceeding of the Digital Imaging Computing: Techniques and Applications DICTA 2005.P.7, ISBN: 0-7695-2467-2.

[8] A. Francois, G. Medioni, " Adaptive Color Background Modeling for Real Time Segmentation of Video Streams", Proc. of International on Imaging Science, Systems, and Technology, pp.227-232,1999.

[9] Peixoto P. Nuno, Cardoso G. Nuno, Cabral M. Jorge, Tavares J. Adriano, Mendes A. Jose "A Segmentation Approach for Object Detection on Highly Dynamic Aquatic Environments", Proc. IEEE Computer Society, Conference on Computer vision and pattern recognition 2009.

[10] How-Lung Eng, Junxian Wang, Alvin H.Kam and Wei-Yan Yau, "Novel region-based modeling for human detection within highly dynamic aquatic environment" In Proc. IEEE Computer Society Conference on Computer vision and pattern recognition, 2004.

[11] A. Francois, G. Medioni, " Adaptive Color Background Modeling for Real-Time Segmentation of Video Streams" Proc, of International on Imaging Science, Systems and Technology,pp.227-232,1999.

[12] C. Ridder, O. Munkelt, H. Kirchner, "Adaptive Background Estimation and Foreground Detection Using Kalman-filtering", Proc. ICRAM'95, pp.193-199, 1995.

[13] Chris Stauffer and W.E.L. Grimson. "Adaptive background mixture models for real-time tracking". In Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, 1999.

[14] Nir Friedman and Stuart Russel. "Image segmentation in video sequence: A probabilistic approach". In Proc. of the Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI), Aug,1-3, 1997.

[15] Rafik Bourezak, Guillaume-Alexandre Bilodeau "Object detection and tracking using iterative division and correlograms," The 3rd Canadian conference on Computer and Robot Vision (CRV'06).

[16] J. Connell, A.W. Senior, A. Hampapur, Y.-L Tian, L Brown, S. Pankanti " Detection and Tracking in the IBM people vision system" IEEE ICME Taiwan June, 2004

[17] I.Haritoglu, D. Harwood and L.S. Davis. "W4: Who? When? Where? What? A Real-Time System for Detecting and Tracking People". In Third Face and Gesture Recognition Conference, Apr., 1998

[18] Ismail Haritaoglu, David Harwood and Larry S. Davis "W4: Real-Time Surveillance of People and Their Activities", IEEE Transactions on pattern analysis and machine intelligence, VOL. 22, NO.8, August 2000.

[19] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell and Alex Paul Pentland. "Pfinder: Real-Time Tracking of the Human Body", in Proc. IEEE Trans. On Pattern Analysis and Machine Intelligence, vol.19, no.7, July, 1997.

[20] Wenmiao Lu and Yap-peng Tan "A camera based system for early detection of drowning incidents" In Proc. IEEE ICIP 2002.

[21] Jae-Chang shim, chitra Dorai, "A Generalized Region Labeling Algorithm for Image Coding, Restoration, and Segmentation". ICIP99, pp.46-50, 1999.

[22] K.V. Mardia. Statistics of directional data. Academic Press, London, 1972.

[23] B.C. Lovell, P.J.Kootsookos, and R.C Williamson. "The circular nature of discrete-time frequency estimates." In IEEE International Conference on ASSP, pages 3369-3372, Toronto, May 1991.

[24] Jae-Chang shim, chitra Dorai, "A Generalized Region Labeling Algorithm for Image Coding, Restoration, and Segmentation". ICIP99, pp.46-50, 1999.