Method for estimation of Background and Foreground segmentation in video surveillance

Freddy Alexander Díaz Gonzáles
Universidad Sergio Arboleda, Bogotá D.C, Colombia, freddy.diaz@correo.usa.edu.co

David Alejandro Arévalo Suárez
Universidad Sergio Arboleda, Bogotá D.C, Colombia, david.arevalo@correo.usa.edu.co

ABSTRACT

The video surveillance systems are widely used today, but it is essential to implement systems that automatically analyze the information captured, with the objective of identifying regular events that may occur in the scene. This work focuses on the estimation of the background and foreground segmentation in a surveillance video. The main contribution is the median filter applied recursively on a temporary popup window, which provides more robustness against noise caused by illumination changes and vibrations in the camera, this without increasing the computational cost.

Keywords: Temporal median, foreground, background, recurrence

1. INTRODUCTION

Today it has increased the use of surveillance systems with video cameras (SSVC), this is partly due to its low cost, ease of installation and maintenance (Pava, 2011) (Kim, 2008). For example, it is estimated that in the UK there are about 4 million cameras, equivalent to a camera for every 14 habitants (Knuth, 2005). However, the information collected is not useful if it's not analyzed and due to the high volume of video captured, their analysis should be done automatically.

The automatic analysis of the video sequences, it has become an active area of research (Plataniotis, 2005) which aims to capture, analyze, identify and alert on unwanted behavior patterns or extracted from videos taken by the SSVC (Ferrando, 2006). SSVC usually installed in certain areas permanently where useful information is sought in these video sequences are dynamic regions, such as pedestrians, abandoned packages and other items not belonging to the monitored area, known as the Foreground (FG) (Mahadevan, 2008). Permanent images in video sequences are called Background (BG) (Parks, 2008).

The most common paradigm for identification of the FG is to have a model of BG. Since a video is a sequence of frames or boxes, the FG will result from the difference between the video frame being analyzed and the model of BG which one has at that moment (Parks, 2008).

This paper proposes a Background Estimation recursive method (BGEM), which is used for segmentation of FG in surveillance videos. The results are validated against an expert system, modifying the values of the main components of BGEM, for analysis of errors in the estimation of BGEM and FG. Additionally it is analyzed the complexity of BGEM according to variations in the model parameters. Section 2 of this paper, we analyze the different methods for estimating BG and segmentation of FG, in Section 3 formulates the model developed, in Section 4 shows the results of tests carried out against the expert system, which are analyzed and presented in Section 5, finally Section 6 concludes the paper and presents future work lines.

2. REVIEW OF METHODS FOR ESTIMATING BACKGROUND PATTERNS IN VIDEOS

Many of the methods used for the analysis of videos, which aim to segment the FG, are based on a priori knowledge of the BGEM, which can be estimated from a sequence of frames of the scene that contains no dynamic elements (Kamijo, 2000) (Gonzalez, 2002). However, not adjust the BGEM to different changes in light and noise generated by vibrations in the camera, causing the increase in false positives in the segmentation of the FG (McHugh, 2009).

One way to reduce false positives in the segmentation of FG is to use methods to update the BGEM with some frequency, so that it fits different types of noise. Within the classification of methods for estimation of BGEM based on updated techniques, there are two broad categories, the first consists of the recursive methods and the second not recursive methods: following a description of each category and the methods used in each.

2.1 NO RECURSIVE METHODS

No recursive methods are those based BGEM estimate in "n" video frames previously stored in a buffer (Toyama, 1999) (Lo, 2001), the estimated models are adaptable to environmental changes and input noises. In the methods non-recursive there are models that calculated the average value to estimate the value of each pixel of the BGEM (Cucchiara, 2003), other methods use the median that is more robust against background noise (Caldera, 2006). Due to the type of frame store that require non-recursive methods, the system which is implemented presents large memory requirements (Maddalena, 2008).

2.2 RECURSIVE METHODS

The recursive methods are those that update the BGEM with input frames (Parks, 2008) (Cheung, 2004). Within these methods are statistical models, for example in (Stauffer, Learning patterns of activity using real-time tracking , 2000) and (Tsai, 2009) analyze the density of the Gaussian distribution behavior of a frame to frame pixel to determine the belonging or not of this to BGEM. Gaussian models are also used to estimate multimodal funds, as presented in (Power, 2002) (Zivkovic, 2006) (Heikkila, 1999), which combine "n" Gaussian models to estimate the "n" BGEM for each pixel channel. More complex statistical models are presented in (Aach, 1995) (Paragios, 2001) (Migdal, 2005) in which are integrated methods like Markov chains and Bayesian networks, in a union based on the general model of Ising.

Statistical recursive methods updated constantly the BGEM, which increases the robustness of the FG segmentation algorithms, due the BGEM fit to variations in lighting changes and other input noises (Marco Cristani, 2010). However, depending on the frequency with which enter a new frame to the model, errors in the estimation can be kept for long periods of time (Maddalena, 2008) and calculations of Gaussian distributions have a high computational cost (McHugh, 2009).

Another variation of a recursive method is the Kernel model, which is typically a Gaussian, this method estimates the probability density function for each pixel called "Kernel Density Estimation" (KDE) (Ahmed, Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance, 2002) (Ahmed, Non-parametric Model for Background Subtraction, 2000). Unlike the Gaussian mixture models (GMM Gaussian Mixture Model in English), the model KDE is a more general approach that does not assume any specific form of the density function (Friedman, 1997) (Stauffer, Adaptive background mixture models for real-time tracking, 1999). This method is widely used for its properties of continuity, differentiability and location. However, the biggest problem with using this technique is its computational cost (Elgammal, 2001).

The feature of update the BGEM with both recursive and non-recursive methods, increases their robustness to noise factors, reducing false positives in the segmentation process of FG. But that same characteristic causes objects not belonging to the bottom of the video, are incurred as part of the fund after remaining static in the scene, during a given video sequence (Caldera, 2006). One way to avoid this effect may be based on the incorporation of a memory parameter in the estimation of BGEM, for each new BGEM generated remember previous models, preventing non-background elements are assumed like part of it.

3. METHOD FORMULATION

The main advantages of the recursive methods that are based on the median filter are the high computational efficiency and robustness to noise (McFarlane, 1995) (Kamijo, 2000) (Gonzalez, 2002) (Marco Cristani, 2010). Additionally, errors in the estimates obtained based on the median BGEM are highly comparable with the results obtained by more complex methods such as those based on GMM (Parks, 2008).

Today most video sequences captured by SSVC frames are stored in color, however, processing for segmentation of the FG can sequences fastest monochrome or grayscale (Jung, 2009). This is because in a gray image, the red, green and blue component of a pixel (RGB color), have the same value (Zhu, 2009), whereby a processing of a picture belonging to a video sequence, it can carry out work only with information from one of the three RGB matrix (Gonzalez, 2002).

The following is a recursive estimation method BGEM and Foreground Segmentation (FGS), based on the median filter. The FGS estimates the BGEM on a limited number of consecutive frames from a video, hereinafter defined as moving temporal window. Additionally, the proposed model provides memory parameters previous BGEM, an analysis to select the range of the threshold used in the segmentation of FG process and determination of the degree of complexity of the model given by the size of the temporary median.

3.1 ESTIMATION MODEL BGM

BGEM is constructed according to the equation:

$$FC_i = \begin{cases} FA * median\left(VC_0 : VC_i\right) + (FM * FC_{i-1}) & \mathrm{i} < x \\ FA * median\left(VC_{i-x} : VC_i\right) + (FM * FC_{i-1}) & \mathrm{i} \geq x \end{cases}$$

Figure 1: Model Equation

Where

$FC_i =$ Background frame of the video it is being estimated.

$FA =$ Gain factor that will have the mobile medium in the current background frame.

$VC_i =$ Video frame entering the estimate of the current background picture.

$FM =$ Gain factor that will have the background frame.

$x =$ Number of frames on which applies the median filter.

$(VC_0 : VC_i) =$ Number of frames that are used to estimate the 'i' first frames when i ≤ x

In the proposed method, the estimate of BGEM consists of weighing between the BGEM calculated recursively applying the median filter ($median\left(VC_{i-x} : VC_i\right)$) and the calculated BGEM immediately preceding (FC_{i-1}.) The proportion between them in the BGEM FC_i is controlled by the constant FA and FM, which must comply FA + FM = 1.

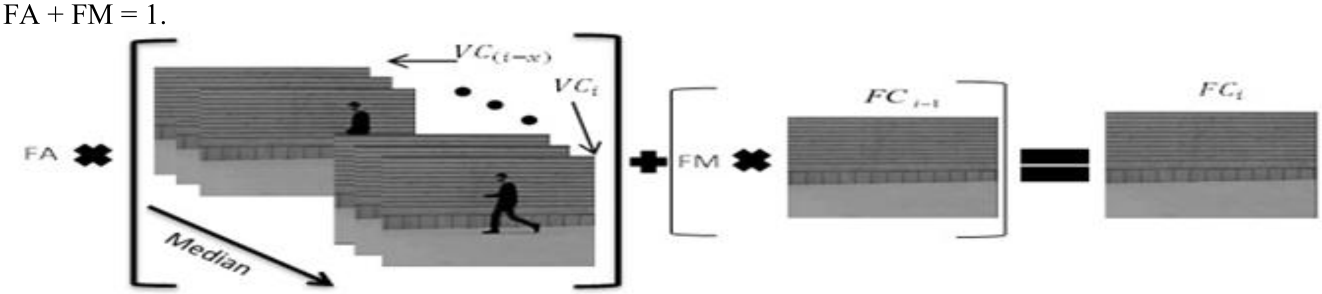

Figure 2: Equation graph model

The main contribution of the method is the inclusion in the calculation of the BGEM, a memory component that reduces the error in estimating the BGEM, caused by non-background elements, which have a low dynamic scene analysis.

3.2 FG SEGMENTATION METHOD

The behavior of a pixel in the sequence of frames that make up the video surveillance is shown in Figure 3. It can be observed in this figure, the pixel value is not constant in each frame, due to induced noise and vibration environment of the capture system (video camera). However, the pixel value is maintained within a range determined, provided that the pixel belongs only to the background of the scene. For example in Figure 3, the pixel value between 0 and 35 frames, varies between 0.4 and 0.45, a priori known that in this range only frame pixel belongs to the background of the scene. Between the frames 36 and 47 the pixel belongs to the FG of the scene, which is reflected in an abrupt change of value.

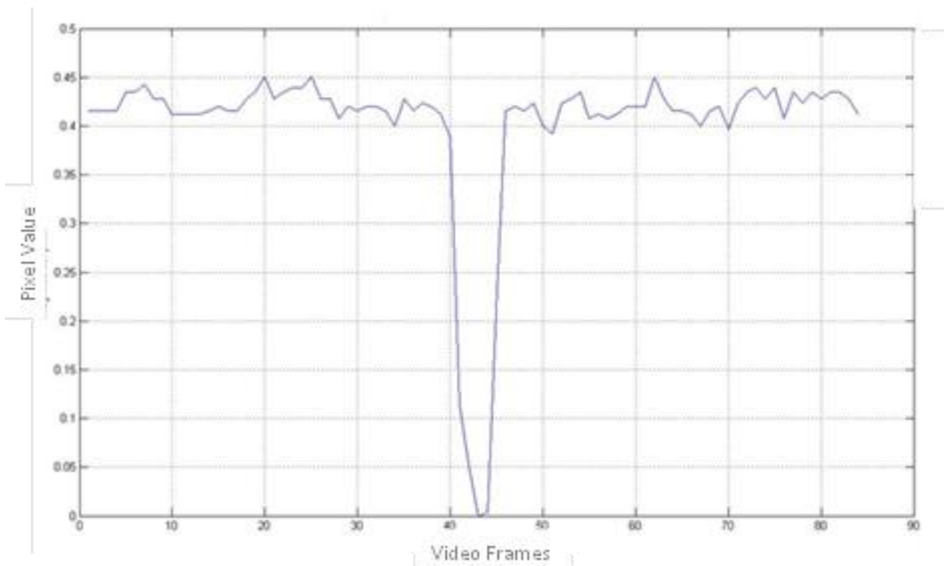

Figure 3: Behavior of a pixel

The foreground segmentation based on a BGEM generated by FGS, which serves as a reference in a single classifier, which determines if the value of a pixel belonging to the new frame in the sequence, is or not, outside a range value. The comparison range of the classifier is bounded by the same pixel value in the BGEM estimate, plus or minus a threshold.

The pixels whose value is not within the threshold are considered as part of FG and the rest are classified as the background of the analyzed scene. The model to segmentation of FG is given by the following equation:

$$PC_i = \left\{ \begin{array}{ll} 1 & (FC_i - \mathrm{r}) < VC_i < (FC_i + \mathrm{r}) \\ 0 & ---- \end{array} \right\}$$

Figure 4: Equation segmentation FG

Where

$PC_i =$ Foreground frame being built

$r =$ Threshold value to determine whether or not belonging to the FG

As shown in equation (figure 4), the value of a pixel in the frame analyzed must stay above of $(FC_i - \mathrm{r})$ but not exceed the value of $(FC_i + \mathrm{r})$, to be identified as part of the bottom, which in this case is classified with a value of 1, in which color values represents white. Values that do not meet the condition of staying in range are classified as FG with a value of 0, which is the color black.

$$\mathrm{A} = \mathrm{var}(VF_0 : VF_x)$$
$$\mathrm{MA} = \mathrm{average}(A)$$
$$\mathrm{MaxMA} = max(MA)$$
$$\mathrm{MMA} = \mathrm{average}(MA)$$
$$\mathrm{r} = \mathrm{MMA} + \mathrm{MaxMA}$$

$$(VF_0 : VF_x) = \text{The first x frames of video test are analyzed to obtain r}$$
$$A = \text{Matrix of variance pixels of the x frames analyzed}$$
$$MA = \text{Vector average of the variance of the pixels in the width of video}$$
$$\mathrm{MaxMA} = \text{The maximum value in the vector MA}$$
$$\mathrm{MMA} = \text{The average of vector MA}$$

The calculation of the threshold r is given initially by the variance of the pixels in x number of frames of video test. A is a matrix of size i x j, as VF. Then calculated the average of A, and the result is stored in a vector MA of size i. The vector MA take out the maximum and average, which when added give the value of r.

3.3 FGS VALIDATION (TESTING)

The FGS model was implemented in Matlab software tool and the surveillance video used for testing has the following features:

- 84 grayscale frames
- Frames of 320 x 160 pixels
- Varying lighting conditions
- Foreground with continuous motion in the video

To validate the FGS established the following conditions:
Contrast FG segmentation provided by the FGS to a specific case against the same segmentation developed by an expert system.
Contrast the BGEM estimate provided by the FGS to a specific case in which there is no current presence of FG, but it was in earlier tables.
To establish the expert system is taken as the analysis frame 67 of the test video sequence. With the aim of rapidly validating the model, are selected from specific areas of the box, as shown in Figure 5.
Figure 5 is the frame 67 of the test video sequence. The scene is changing lighting and camera vibration, noise inducing SSVC. The behavior of a test video pixel is represented in Figure 3. At the scene of the dynamic element analysis moves from right to left, the red box in Figure 5 represents an area that should only be classified as the background of the video, but which has passed the dynamic element, the box has a dimension of 97x46 pixels.

Figure 5: Regions of interest

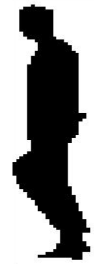

Figure 6: Manual Segmentation

The blue outline of the dynamic element is segmented manually in a box of 97x46 pixels. The expert system is set to the red box in Figure 5, which is just over the scene, and manual segmentation of the dynamic element as the FG stage.

Figure 6 shows the 97x46 pixel area on which FG segmentation manual was done of the frame 67, taken as part of the expert system for validation of the FGS.

For testing the FGS in the estimation of BGEM and segmentation of FG, were varied the following parameters of the model:

FA, FM and x from equation of Figure 1

FA value begins with a value of 0.1, increasing in 0.1, until the value of 0.9. Since it must keep FA + FM = 1, the value of FM starts and finishes at 0.9 to 0.1.

The parameter x that gives the size of the moving temporal window over which the filter applies median concurrently began with value of 3, is increased by 1, reaching a value of 15 to analyze the behavior of the model in different sizes moving time window.

4. RESULTS

After testing the model with different values for the parameters FA, FM and x, as mentioned previously, data are taken of the generated errors in the estimation of BGEM and segmentation of FG, we present the error behavior in Figures 7 and 8.

Figure 8 shows the value of the error in estimating the background generated by FGS, with respect to window size and the gain factor FA.

Figure 7 shows the value of the error in the segmentation of FG generated by FGS, with respect to window size and the gain factor FA.

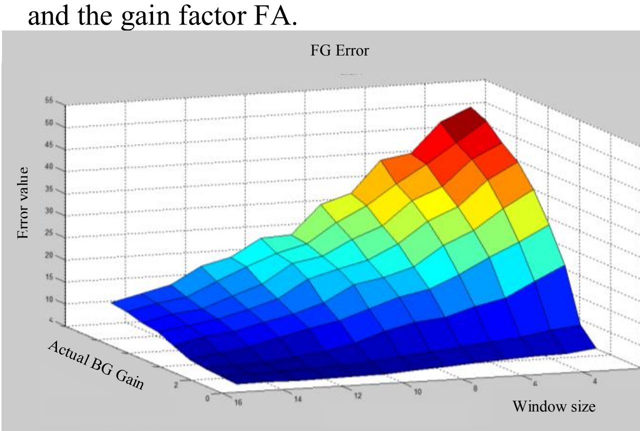

Figure 7: Error FG model

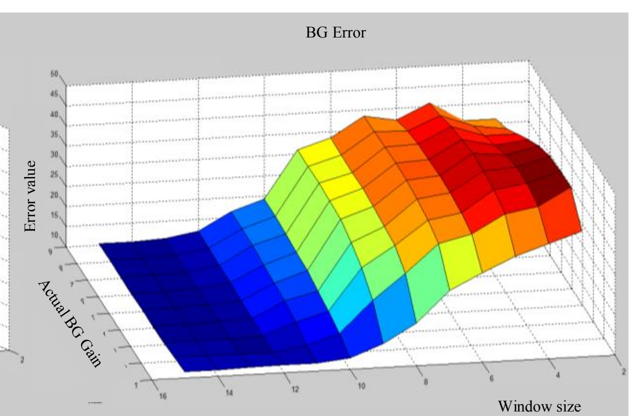

Figure 8: Error BG model

In the graph 7:

- The error in the segmentation of FG decreases with increasing size of the parameter x.
- The greatest error in segmentation is presented with FG value of x equal to 3, and FA value of 0.9.
- In FG segmentation error increases if FA grows.
- The minor errors in the segmentation of the FG are presented with FA of 0.1 regardless of the size of x.

In Figure 8:

- The error in estimating the BGEM decreases with increasing size of the parameter x.
- BGEM in the estimation of the biggest mistakes occur with values of x between 3 and 6, with FA values between 0.2 and 0.8.
- The error in estimating the BGEM is not affected by varying FA thoughtfully.
- BGEM in the estimation of the errors occur with lower values of x between 12 and 15, regardless of the value of FA.

FGS model complexity depends only on the parameter x, which is the number of frames in the moving temporal window for any video sequence. This is because the model is based on BGEM median filter. In implementing this filter arrangement should be organized and take the value of the middle position, where the computational cost is focused on the organization of the array and is given by f * log (f) (Knuth, 2005), where f is the size arrangement for the FGS method f = x.
Behavior model complexity BGEM in function of the variation of the parameter x is shown in Figure 9, where the complexity increases if x increases.

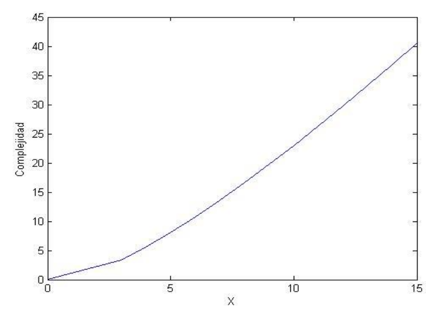

Figure 9: Complexity of the method

5. ANALYSIS OF RESULTS

The model developed to estimate the BGEM and segment the FG, it works well for values of x greater than 9, which are minor errors in the tests, since the median filter to have more data to compare values can be removed easily extremes such as non-background pixels.
If the FA parameter value increases, FG segmentation error grows, since it is being amplified, or giving more weight to the values obtained in the median, thus, the pixels to be most of the background, some achieve to exceed the threshold after FA gain them, and are classified as part of FG. The pixels that should be classified as part of BGEM, but included in the FG segmentation are errors of false positives.
Selecting the right value for the model parameters r and FA in FGS, allows obtaining the best classification of the pixels in the segmentation of FG avoiding false positives.

Given that increasing the value of x increases the complexity of the proposed method "FGS", you must set a value of x to decrease the computational cost and generate a large amount of errors in estimating the BGEM and segmentation of the FG. With values of x between 9 and 15 the model error is smaller.
In the modification of the FA parameter, values between 0.1 and 0.3 generated little error in estimating the BGEM and segmentation of the FG.
The combination of the parameters x and FA for the minimum error in the method FGS are considered values of x between 9 and 11, and FA between 0.1 and 0.2.
The FGS model had the lowest error when working with x values between 9 and 11, and the combination of FA between 0.1 and 0.2.

6. CONCLUSIONS

The implementation of the temporal median filter concurrently on a mobile temporal window, it reduced the error in the estimation of BGEM. Additionally, it decreased the computational cost and memory requirements by limiting the number of frames incorporated in the method for estimating the BGEM.
The proposed model performs a segmentation of the foreground with a very low number of false positives due to the low error in estimating the BGEM and the range used in the proposed single classifier.
The addition of a memory factor in the estimation of BGEM process, it provides a robust method FGS against low dynamic elements that appear on the scene, strengthening them in the foreground segmentation.
For future work is proposed to process video online, with the FGS method implementation in a language of low level, for use in embedded systems, to develop commercial applications of automated surveillance systems at low cost.

REFERENCES

Aach, T. (1995). Bayesian algorithms for adaptive change detection in image sequences using Markov random fields. *Signal Process.,, 7*, 147-160.

Abbott, R. G. (2009). Multiple Target Tracking with Lazy Background Subtraction and. *Machine Vision and Applications*, 93-101.

Ahmed, E. (2000). Non-parametric Model for Background Subtraction. *Proceedings*, 751-767.

Ahmed, E. (2002). Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance. *Proc. IEEE, 90*, 1151-1163.

Bayona, Á. (2009). Comparative evaluation of stationary foreground object detection algorithms. *Advanced Video and Signal Based Surveillance, Sixth IEEE International Conference on*, (pp. 25-30). Genova.

Benezeth, Y. (2008). Review and Evaluation of Commonly-Implemented Background. *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, (pp. 1-4). Tampa, FL.

Benezeth, Y. (2010). Comparative study of background subtraction algorithms. *J. Eectron. Imaging*.

Caldera, S. (2006). Reliable background suppression for complex scenes. *Video surveillance and sensor networks*, 211-214.

Chae, H.-U. (2008). Identification of a Human using Accorded Blobs on the Varied Region from. *Control, Automation and Systems, 2008. ICCAS 2008. International Conference on*, (pp. 1887-1891). Seoul.

Cheung, S.-C. S. (2004). Robust techniques for background subtraction in urban traffic video. *Visual Communications and Image Processing, 5308*, 881-892.

Cucchiara, R. (2003). Detecting moving objects, ghosts, and shadows in video streams . *Pattern Analysis and Machine Intelligence, IEEE Transactions on , 25*(10), 1337-1342.

Davis, J. W. (2007). Background-subtraction using contour-based fusion. *Computer Vision and Image Understanding*.

Elgammal, A. (2001). Efficient non-parametric adaptive color modeling using fast Gauss transform. *Computer Vision and Pattern Recognition IEEE Computer Society Conference on*, 563-570.

Ferrando, S. (2006). Classification of Unattended and Stolen Objects in Video-Surveillance System. *Video and Signal Based Surveillance.* Sydney, Australia.

Friedman, N. (1997). Image segmentation in video sequences: a probabilistic approach. *13th Conference on Uncertainty in Artificial Intelligence*, 175-181.

Gonzalez, R. C. (2002). *Digital Image Processing.* Prentice Hall.

Heikkila, J. (1999). A real-time system for monitoring of cyclists and pedestrians. *Visual Surveillance, 1999. Second IEEE Workshop on*, 74-81.

Herrero, S. (2009). Advanced Concepts for Intelligent Vision Systems. *Lecture Notes in Compter Science*.

Hung, M.-H. (2010). Speed Up Temporal Median Filter for Background Subtraction. *Pervasive Computing Signal Processing and Applications (PCSPA), 2010 First International Conference on*, (pp. 297-300). Harbin.

Jung, C. R. (2009). Efficient Background Subtraction and Shadow Removal. *Multimedia, IEEE Transactions on*, 571-577.

Kamijo, S. (2000). Occlusion robust tracking utilizing spatio-temporal markov random field model. *ICPR, 1*, 1140.

Kim, Z. (2008). Real Time Object Tracking based on Dynamic Feature Grouping with. *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* , (pp. 1-8). Anchorage, AK.

Knuth, D. E. (2005). *Art of Computer Programing.* United States: Addison-Wesley .

Lo, B. (2001). Automatic congestion detection system for underground platforms. *Intelligent Multimedia, Video and Speech Processing, 2001. Proceedings of 2001 International Symposium on* , 158-161.

Maddalena, L. (2008). A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 1168-1177.

Mahadevan, V. (2008). Background Subtraction in Highly Dynamic Scenes. *Conf. on Computer Vision and Pattern Recognition* (p. 1). Anchorage, AK: IEEE.

Marco Cristani. (2010). Background Subtraction for AutomatedMultisensor Surveillance:. *EURASIP Journal on Advances in Signal Processing*.

McFarlane, N. J. (1995). Segmentation and tracking of piglets in images. *Machine Vision and, 8*(3), 187–193.

McHugh, J. M. (2009). Foreground-Adaptive Background Subtraction. *Signal Processing Letters, IEEE*, 390-393.

Migdal, J. (2005). Background Subtraction Using Markov Thresholds . *Motion and Video Computing, 2005. WACV/MOTIONS '05 Volume 2. IEEE Workshop on* , 58-65.

Monnet, A. (2003). Background modeling and subtraction of dynamic scenes. *CVPR*.

Panahi, S. (2008). Ecaluation of Background Subtraction Methods. *Techniques and Applications*, 357-364.

Paragios, N. (2001). A MRF-based approach for real-time subway monitoring. *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on , 1*, 1034-1040.

Parks, D. H. (2008). Evaluation of Background Subtraction Algorithms with Post-processing. *Advanced Video and Signal Based Surveillance, 2008. AVSS '08. IEEE Fifth International Conference on*, (pp. 192-199). Santa Fe, NM.

Pava, D. F. (2011). Object Detection and Motion Analysis in a Low Resolution 3-D Model. *Ninth LACCEI Latin American and Caribbean Conference (LACCEI'2011), Engineering for a Smart Planet, Innovation, Information Technology and Computational Tools for Sustainable Development.* Medellín, Colombia.

Plataniotis, K. (2005). Visual-centric Surveillance Networks and Services. *IEEE Signal Processing Magazine, 22*(2), 12-15.

Power, P. (2002). Understanding background mixture models for foreground segmentation. *Proc. of the Image and Vision Computing*, 266-271.

Sigari, M. H. (2008). Fuzzy Running Average and Fuzzy Background Subtraction:. *IJCSNS International Journal of Computer Science and Network Security*, 138-143.

Stauffer, C. (1999). Adaptive background mixture models for real-time tracking. *IEEE Computer Society Conference on Computer Vision and, 2*, 246–252.

Stauffer, C. (2000). Learning patterns of activity using real-time tracking . *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22*(8), 747-757.

Toyama, K. (1999). Wallflower: principles and practice of background maintenance. *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on , 1*, 255-261.

Tsai, D.-M. (2009). Independent Component Analysis-Based Background. *IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 18, NO. 1*, 158-167.

Zhou, Q. (2001). Tracking and classifying moving objects from video. *IEEE Proc. PETS Workshop*.

Zhu, F. (2009). A Video-based Traffic Congestion Monitoring System Using Adaptive Background Subtraction. *Electronic Commerce and Security, 2009. ISECS '09. Second International Symposium on*, (pp. 73-77). Nanchang.

Zivkovic, Z. (2006). Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters, 27*(7), 773-780.

Zivkovic., Z. (2004). Improved adaptive Gaussian mixture model for background subtraction. *ICPR*.

Authorization and Disclaimer

Authors authorize LACCEI to publish the paper in the conference proceedings. Neither LACCEI nor the editors are responsible either for the content or for the implications of what is expressed in the paper.