

Anomaly Detection in Surveillance Video using Color Modeling

M. Gangadharappa
Ambedkar Institute of
Advanced Communication
Technologies & Research,
Delhi, INDIA

Pooja Goel
Indira Gandhi Institute of
Technology, Kashmiri Gate
Delhi, INDIA

Rajiv Kapoor
Delhi Technological University
Delhi, INDIA

ABSTRACT

The primary goal of this paper propose an algorithm for automatic detection of abnormal events in video surveillance scenarios. We specifically focus our attention on the event of object dropping in public places such as railway stations and airports etc. We look into how to distinguish events in surveillance video, and further what is a remarkable event. Analyzing surveillance data, without the knowledge of when and where or even if an interesting event has occurred which often takes place, is very time consuming labour. In this kind of analysis we are interested in extraordinary events, something that deviates from the normal.

Keywords

Abnormal events, surveillance videos, object tracking, feature- extraction and feature- analysis.

1. INTRODUCTION

In this paper, we consider the problem of detecting anomalous or abnormal event of object dropping in public places such as railway stations and airports etc. The need for heightened security is unfortunately becoming more prevalent in today's world. While video cameras are intended to capture the images of possible offenses, the majority of the time, it is done but a security person who spends his or her most of the time staring into a series of uneventful monotonous images. Since new events occur rarely, it is extremely difficult for a security person to remain attentive at all times. Thus, our work removes the onus of detecting anomalous situations from the security person; and rather, places it on the video surveillance system.

In this paper, we have used YUV or YCbCr color model which differs from RGB model. It comprises the luminance (Y) and two color difference (U, V) components. The principal advantage of the YUV model in image processing is decoupling of luminance and color information.

In the field of video processing there are several recently proposed methods for the detection of abnormal events in video surveillance scenarios. The methods can be grouped according to the dimensions of the sampling support like pixel, region of interest(ROI), Frame based etc., the features and the mathematical modeling technique used (see Table 1). The sampling support defines the input that is used to find a decision about the abnormality. The smallest sampling

support is a single pixel. Ermis et al. [1] learn the statistical properties of each pixel as an independent sensor and detect abnormality based on its (foreground/ background) status. Frame-based methods model normal behaviours of the full scene within the field of view of a camera. The scene descriptor information is commonly extracted based on moving objects detection (blobs), and the models are trained to detect unusual events. Li et al. [2] use document analysis techniques together with object detection to find abnormal actions taking place in large segmented region of interest in a scene.

TABLE 1

Abnormality detection using low-level features

Sampling support	Features	Modeling	Reference
Pixel	Object detection	Density estimation	Ermis et al. [1]
Region	Object detection	pLSA	Li et al. [2]
		Probability distribution	Adam et al. [3]
Frame	Motion vectors	Graph co-clustering	Zhong et al. [4]
		Multiobservation HMM	Zhang et al. [5]

Different authors have adopted different methodology for abnormality detection based on merits and demerits (see Table 2). Xiang et al. [6] use MOHMM (Multi Observed Hidden Markov Model) and FBR (Forward Backward Relevance). Xiang et al. [7], also uses Incremental and adaptive mode learning, fully supervised, LRT (Likelihood Ratio Test), EM (Expectation Maximization). Chen ChangeLoy et al. [8] decomposes a complex behaviour pattern according to its temporal characteristics or spatial-temporal visual contexts, modeled using a Cascade of Dynamic Bayesian Networks (CasDBNs).

TABLE 2
Methodologies used for abnormality detection in video

Methodology	Parameters	Merits
MOHMM, FBR, Propose recently-spectral clustering algorithm. Xiang et al. [6]	Based on break points on a high dimensional video - content trajectory. 7 - D feature space using GMM. GMM is estimated using EM.	Requires less parameters compared to conventional HMM. Run in real time compared to sliding window.
Incremental and adaptive mode learning, Fully supervised, LRT (Likelihood ratio Test), EM. Xiang et al. [7].	Affinity matrix, 7dimensional feature vector, GMM with BIC is used to classify event in classes.	Computationally efficient, Suitable for real-time applications, Fully unsupervised.
Uses low level features. Cooccurrences matrix is used as a potential function in Markov random field to detect moving object whose behavior differs from the ones observed during the training phase Yannick et al. [9].	Cooccurrence matrix is used. Markov random field distribution accounts for speed, direction, average size of object without any higher level intervention.	Based on scenarios through spatio temporal models. Based on-spatial correlation. Simple Abnormal behavior detection first followed possible by object extraction and tracking.
Present a non- linear subspace learning detector to address the abnormality detection problem. Patterns in the selected features are normally modeled by probabilistic method such as HMM Ioannis et al. [10].	Motion vectors extracted over a region of interest (ROI) as features and a non-linear, graph-based manifold learning algorithm coupled with a supervised novelty classifier.	Detects abnormal events without prior knowledge of normal patterns. Does not require high-level information filtering steps (e.g., object tracking). Accounts for the non-linear correlation of the features is generic and can be used with a variety of low-level features.

Yannick et al. [9] uses low level features like Co-occurrences matrix is used as a potential function in Markov random field

to detect moving object whose behaviour differs from the ones observed during the training phase. Ioannis et al. [10] present a non- linear subspace learning detector to address the abnormality detection problem. Patterns in the selected features spaces are normally modeled by probabilistic method such as HMM. Graph based non linear dimensionality reduction is used. Tziakos and Cavallaro [11] use On-line abnormality detection Local abnormality modeling, PCA (Principle Component Analysis), LPP (locality preserving projections). Novelty detection consist of GMM, EM, and BIC. Wei Wang et al. [12] uses PCA to reduce dimensions of the histogram matrix, information of frames differences is used to describe the motion change information, and the information twice frame difference is used to describe the change rate. Ernesto L. Andrade et al. [13] uses Background modeling and optical flow computation, also unsupervised feature extraction. Spectral clustering, HMM.

This paper is organized as follows. Section 2 will provide an overview of the proposed approach. Section 3 will provide video pre-processing along with foreground segmentation methods. Sections 4 and 5 will cover object detection methodology and further discusses feature extraction, analysis. Finally, implementation is described in section 6; results are discussed in section 7, preceding the conclusions presented in section 8.

2. OVERVIEW OF METHODOLOGY

In the field of video processing there are several recently proposed methods for the detection of abnormal events in video surveillance scenarios. Most systems today consist of the steps presented in figure 1.

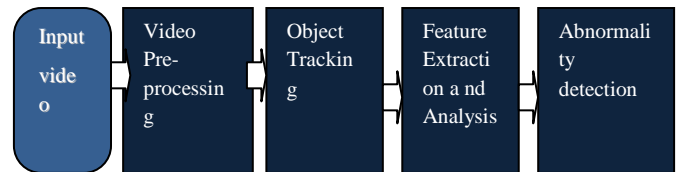


Figure 1: General system architecture for abnormal event detection.

To develop our algorithm, we used sample video file taken at some area as shown in figure 2. This is a realistic representation of the pedestrian traffic during slow commuting times. We also load corresponding background image taken when the area was empty. We apply the detection and tracking algorithms to each frame of the video. Our algorithm is doing the differencing between frames. The difference is done in the YCbCr color space, a typical video signal color space, to ensure that objects are correctly identified both by lightness (in the greyscale Y channel) and color (in the Cb and Cr channels). From this difference a threshold is used to change the image to black and white. After the image is converted to black and white, noise is removed using morphological operations and then objects are identified using bwlable and the properties for the objects are obtained using region -props like area, centroid, persistence etc.

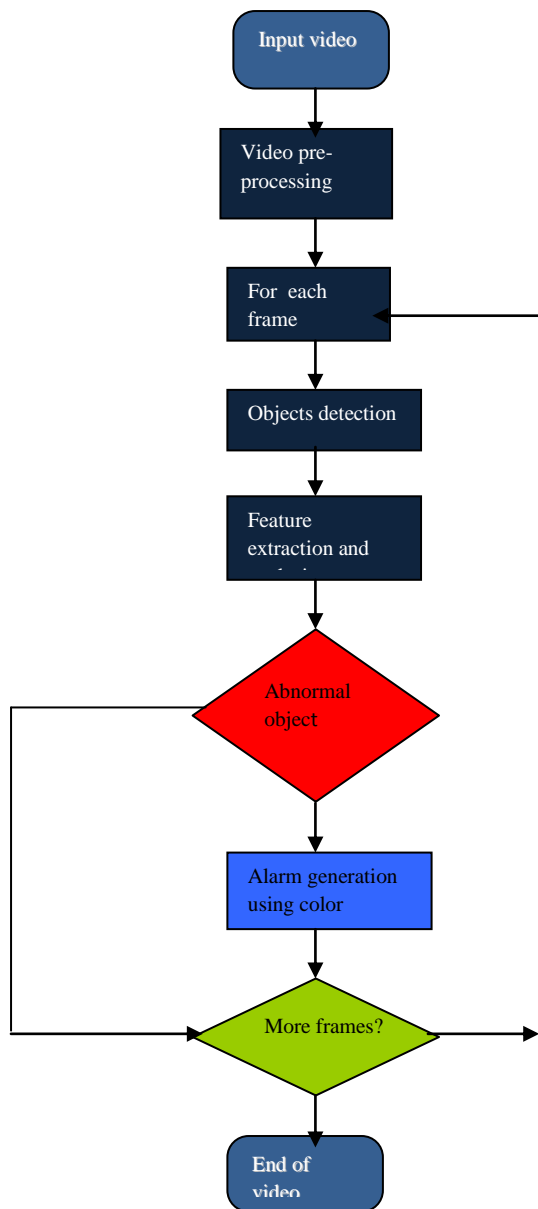


Figure 2: Our Framework

The approach is to first find all objects in the video, then to identify objects that have been stationary for a specified length of time. Finally, We add rectangles for current objects and overlay blue on objects that have been abandoned and we have generated sound vector of tone frequency 50 Hz.

3. VIDEO PRE-PROCESSING

Segmentation is the process of dividing image into different regions. Segmentation shows the objects and edges in an image. Each pixel in the region has some similar characteristics like color, intensity level, etc. The objective is to automatically segment a continuous video sequence V into N video segments i.e. from V_1 to V_N such that ideally each segment contains a single behaviour pattern. The n th video segment V_n consists of T_n image frames. Depending on the nature of the video sequence, various segmentation approaches can be adopted. Since our focus is on surveillance video, the generally used segmentation based on shot change detection approach is not appropriate. In a not-too-busy scenario, there are often non-activity gaps between two consecutive behaviour patterns which can be utilized for activity segmentation. In the case where obvious non-activity

gaps are not available, an on-line segmentation algorithm can be adopted. Alternatively, the video can be simply sliced into overlapping segments with fixed time duration.

3.1 Background Estimation

To be able to track any moving objects in a video, each frame must be subtracted from the background. The estimated background considered for our work is shown in figure 3. This returns an object as the foreground.



Figure 3: Background

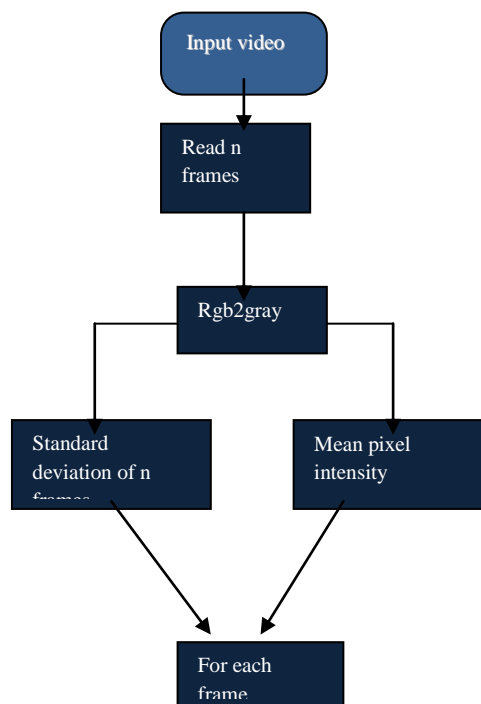


Figure 4: Background estimation

The output of the background estimation is based on two variables; which are taken as mean pixel intensity and standard deviation, both of which come from the set of frames given by the user as input as depicted in figure 4. The result is a single frame (image) that is representative of the set of frames given as input by the user. The background-image is derived by taking the mean value of all pixels at each position in the set of frames. All mean values of each pixel are stored in a new separate matrix. This image will now be considered as the background of the scene, and used to extract the foreground objects of the future frames. According to [14] the most basic adaptive technique is keeping a long-term average of the images in the video sequence which is subsequently and continuously updated. An example of this would be the equation presented in [14]:

$$B(x, y, t) = \frac{1}{t} \sum_{t'=1}^t I(x, y, t') \quad (1)$$

Where $I(x,y,t')$ is the current pixel (x,y,t) value for pixel at time t' and B is the current BM(background model). According to [14], a more incremental version of this equation, in which the current pixel value contributes only a certain portion of the BM, is shown on the following page:

$$B(x, y, t) = \frac{t-1}{t} B(x, y, t-1) + \frac{1}{t} I(x, y, t) \quad (2)$$

The major downfall of this type of incremental equation is the need for storage space to keep the previous $t-1$ pixel values. According to [14] an alternative, exponential version of the long-term average would eliminate those storage requirements and would allow the most recent pixel value to have an α weight, the forgetting constant, in calculating the BM. Jacobs and Pless [15] provide a discussion on temporal decomposition using the exponential equation. The exponential equation is shown below:

$$B(x, y, t) = (1 - \alpha)B(x, y, t-1) + \alpha I(x, y, t) \quad (3)$$

According to Friedman et al. [14] using exponential forgetting to minimize the effect of illumination variation is equivalent to using a Kalman filter [16]. These are also updated accordingly when needed.

4. OBJECT DETECTION

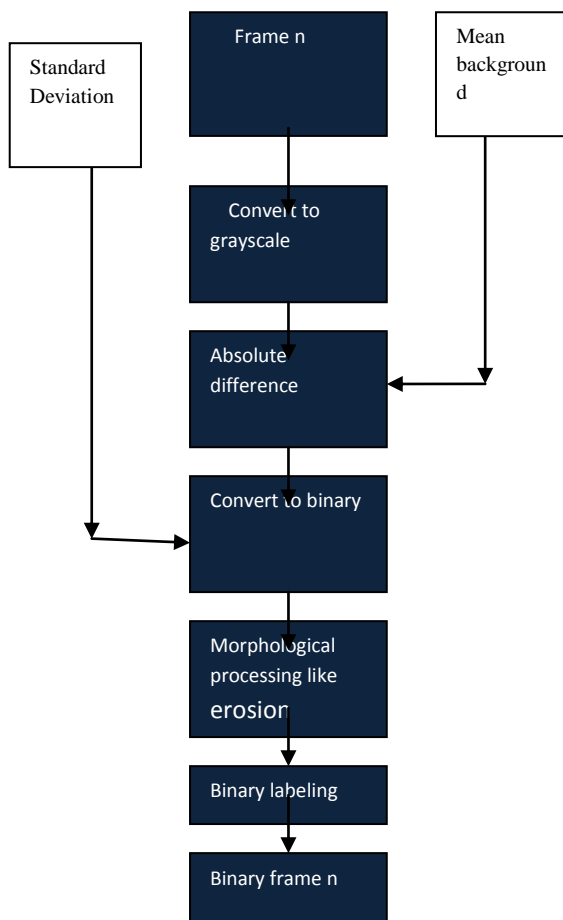


Figure 5: Object detection

To detect any moving objects in the scene, each frame is subtracted from the mean background image. This results in a

foreground image containing one or more objects. This is converted into binary, making the background pixels black and the foreground pixels, scene objects, white. This is done by using the standard deviation and a confidence-level as a threshold. Further each frame goes through a morphological filtering step such as erosion and dilation, which removes small objects in the scene, which are regarded as noise elements. The objects are identified based upon pixel connected component analysis. Additionally any blobs that are not filled, due to bad segmentation are closed. The flow of the algorithm is given in figure no 5.

5. FEATURE EXTRACTION AND ANALYSIS

Feature Extraction plays a major role to detect the moving objects in sequence of frames. Every object has a special feature like color or shape. In a sequence of frames, any one of the feature is used to detect the objects in the frame.

5.1 Bounding Box with Color Feature

If the segmentation is performed using frame difference, the residual image is visualized with rectangular bounding box with the dimensions of the object produced from residual image. For a given image, a scan is performed where the intensity values of the image is more then limit (depends on the assigned value, for accurate assign maximum). In this features are extracted by colors and here the intensity value describes the color. The pixel values from the first hit of the intensity values from top, bottom, left and right are stored. By using this dimension values a rectangular bounding box is plotted within the limits.

5.2 Blob Analysis

Blob analysis is to analyze pixels of image's connectivity which is known as Blob. Blob analysis will have some major jobs to complete. Firstly, Statistic of every blob is calculated. Secondly, pick-up blob's information which is mainly refers to the geometric characteristics, including point of borderline, area, centroid, perimeter and so on. The region-of-interest has to be initialized only one-time at the first frame. Without any further user interaction the entire sequence can be analyzed, which returns the blobs in every frame. Steps are as follows: Convert the resulting grayscale image into a binary image. Then label all the connected components in the image. We set all properties for each labeled region. Last we go for loop to bound the colored objects in a rectangular box.

5.3 Color Models

The purpose of a color model is to facilitate the specification of colors in some standard generally accepted way. Each industry that uses color employs the most suitable color model. For example, the RGB color model is used in computer graphics, YUV or YCbCr are used in video systems. We have used YUV or YCbCr color model.

5.3.1 YUV Color Model

The YUV color model is the basic color model used in analogue color TV broadcasting. Initially YUV is the re-coding of RGB for transmission efficiency (minimizing bandwidth) and for downward compatibility with black-and white television. The YUV color space is "derived" from the RGB space. It comprises the luminance (Y) and two color difference (U, V) components. The luminance can be computed as a weighted sum of red, green and blue components; the color difference, or chrominance, components are formed by subtracting luminance from blue and from red. The principal advantage of the YUV model in

image processing is decoupling of luminance and color information. The importance of this decoupling is that the luminance component of an image can be processed without affecting its color component as shown in figure 6. The Intel IPP functions use the following basic equation to convert between gamma-corrected R'G'B' and Y'U'V' models:

$$Y' = 0.299 * R' + 0.587 * G' + 0.114 * B' \quad (4)$$

$$U' = -0.147 * R' - 0.289 * G' + 0.436 * B' = 0.492 * (B' - Y') \quad (5)$$

$$V' = 0.615 * R' - 0.515 * G' - 0.100 * B' = 0.877 * (R' - Y') \quad (6)$$

$$R' = Y' + 1.140 * V' \quad (7)$$

$$G' = Y' - 0.394 * U' - 0.581 * V' \quad (8)$$

$$B' = Y' + 2.032 * U' \quad (9)$$

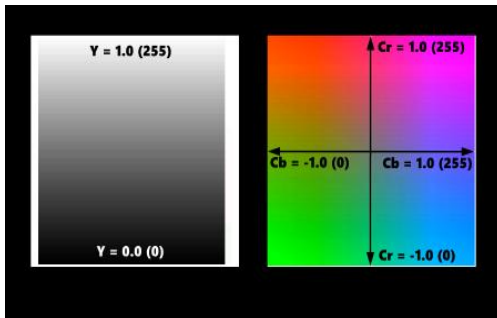


Figure 6: YUV color components

5.4 Alarm Generation

For alarm generation we add rectangles for current objects and overlay blue on objects that have been abandoned. Any valid RGB color falls somewhere within this color space. For example, pure magenta is 100% blue, 100% red, and 0% green. [1 0 1] as shown in figure 7. Also we have generated sound of tone frequency 50 Hz during abnormality detection.

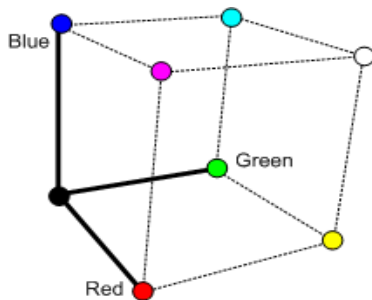


Figure 7

6. IMPLEMENTATION OF ALGORITHM

To develop our algorithm, we used sample video file taken at some area. This is a realistic representation of the pedestrian traffic during slow commuting times. We also load corresponding background image taken when the area was empty. We apply the detection and tracking algorithms to each frame of the video.

Our algorithm is doing the differencing between frames. The difference is done in the YCbCr color space, a typical video signal color space, to ensure that objects are correctly identified both by lightness (in the grayscale Y channel) and color (in the Cb and Cr channels). From this difference a threshold is used to change the image to black and white. After the image is converted to black and white, noise is

removed using morphological operations and then objects are identified using connected component analysis and the properties for the objects are obtained using region -props. We have also used Movavi video converter11 to remove video compression.

7. RESULTS

To test the validity of proposed approach, we experiment on video consist of 781 frames. Figure8 shows some of the frames. Frame no labeled as 282 shows that one object i.e. a person is being tracked with a yellow boundary. Frame 324 shows the same person carrying another object (i.e. a purse) in the hand. In frame 587, person keeping the purse (object) on the floor. In frame 625, purse is being tracked with a yellow boundary. If purse remains stationary for a few frames as per desired threshold for a no of frames then it is considered as abandoned object and overlay with blue color. Side by side alarm is also generated with a tone frequency of 50 Hz. Frame 740 shows two objects, hand of the person with a yellow boundary and abandoned object with overlay of blue color. The results are obtained on the following parameters of the video that we have considered-

VideoCompression: 'none'
Quality: 4.2950e+007
NumColormapEntries: 0
AudioFormat: 'Format # 0x55'
Audio Rate: 44100
NumAudioChannels: 2

Frame #282 # Objects tracked = 1



Frame #324 # Objects tracked = 1





Figure 8. Examples of applying algorithm on our video detecting person with yellow boundary as well as abandoned object in blue color

8. CONCLUSION

The algorithms presented in this paper represent a surveillance video analysis system applied for the task of abnormal event detection of object dropping in public places such as airports and train stations etc. Foreground-segmentation, foreground enhancement, object tracking, feature extraction, and object classification have been applied to a database of videos.

Future work may explore the use of other image features such as wavelet coefficients, windowed averages of object growth, and windowed averages of velocity measurements. It may be beneficial to explore the use of types of classifiers including neural networks.

ACKNOWLEDGEMENTS

Authors are grateful to the All India Council of Technical Education for sponsoring the project and appreciate its contribution by funding the whole project.

REFERENCES

- [1] E.B. Ermiş, V. Saligrama, P.-M. Jodoin, J. Konrad, "Abnormal behavior detection and behavior matching for networked cameras", in: Proceedings of the ACM/ IEEE International Conference on Distributed Smart Cameras (IEEE), New York, 2008, pp. 1–10.
- [2] J. Li, S. Gong, T. Xiang, "Global behaviour inference using probabilistic latent semantic analysis", in: Proceedings of the British Machine Vision Conference, BMVA, Malvern, 2008
- [3] A. Adam, E. Rivlin, I. Shimshoni, D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors", IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (3) (2008) 555–560.
- [4] H. Zhong, J. Shi, M. Visontai, "Detecting unusual activity in video", in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, IEEE, New York, 2004, pp. 819–826.
- [5] Zhang, D. Gatica-Perez, S. Bengio, I. McCowan, "Semi-supervised adapted HMMs for unusual event detection", in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, IEEE, New York, 2005, pp. 611–618.
- [6] Tao Xiang and Shaogang Gong, "Activity based surveillance video content modeling," Pattern Recognition 41(2008) 2309-2326.
- [7] Tao Xiang, Shaogang Gong, "Incremental and adaptive abnormal behaviour detection", Computer Vision and Image Understanding 111 (2008) 59–73.
- [8] Chen ChangLoy, TaoXiang, ShaogangGong, "Detecting and discriminating behavioural anomalies", Pattern Recognition 44 (2011) 117–132.
- [9] Yannick Benezeth, Pierre-Marc Jodoin, Venkatesh - Saligrama, "Abnormality detection using low level cooccurring events", Pattern Recognition 32(2011)423-431
- [10] Ioannis Tziakos, AndreaCavallaro, Li-QunXu, "Event monitoring via local motion abnormality detection in non-linear subspace", Neuro-computing, 73 (2010) 1881–1891.
- [11] Ioannis Tziakos and Andrea Cavallaro, "Local Abnormality Detection in Video Using Subspace Learning", 2010 IEEE, DOI 10.1109/AVSS.2010.70
- [12] Wei Wang, Peng Zhang, Runsheng Wang, "Abnormal Video Sections Detection Based on Inter-Frames Information", IEEE DOI 10.1109/MUE.2009.93.
- [13] Ernesto L. Andrade, Scott Blunsden2 and Robert B. Fisher, "Modeling Crowd Scenes for Event Detection", (ICPR'06),0-7695-2521-0/06,IEEE.
- [14] Friedman, N. and Russell, S. 1997. Image segmentation in video sequences: a probabilistic approach. In Proc. 13th Conf. Uncertainty in Artificial Intelligence, (Brown University, Providence, Rhode Island, USA, August 1-3, 1997). Morgan Kaufmann, San Francisco, CA, 175 - 181.
- [15] Jacobs, N. and Pless, R. 2006. Real-time constant memory visual summaries for surveillance. In Proceedings of the 4th ACM international Workshop on Video Surveillance and Sensor Networks (Santa Barbara, California, USA, October 27 - 27, 2006). VSSN '06. ACM Press, New York, NY, 155-160.
- [16] Welch, G. and Bishop, G. 1995 An Introduction to the Kalman Filter. Technical Report. UMI Order Number: TR95-041., University of North Carolina at Chapel Hill