Journal of Theoretical and Applied Information Technology
15th June 2018. Vol.96. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A NOVEL APPROACH FOR CROWD BEHAVIOR REPRESENTATION: NORMAL AND ABNORMAL EVENT DETECTION

[1]SHERIF EL-ETRIBY, [2,3]MAHMOUD ELMEZAIN, [4]MOEIZ MIRAOUI

[1]Computer Science Department, Faculty of Computers and Information, Menoufia University, Egypt

[2]Computer Science Division, Faculty of Science, Tanta University, Tanta, Egypt

[3]Faculty of Science and Computer Engineering, Taibah University, Yanbu, KSA

[4]High Institute of Science and Technology University of Gafsa, Tunisia

E-mail: [1]El_etriby100 @yahoo.com, [2]Mahmoud.Elmezain@tuscs.com

ABSTRACT

The concept of crowd refers to the gathering many people in one place, such as train stations, airports and subways, as well as gatherings sports, religious special crowds Hajj and Umrah became highly congested. In this paper, a novel approach is investigated to the crowd behaviors of individual using discriminative models. The novelty of the proposed approach can be described in three aspects. First, we sectioned video segments into spatio-temporal flow-blocks which allow the marginalization of arbitrarily dense flow field. Second, the observed flow field in each flow-block is treated as 2D distribution of samples and mixtures of Gaussian is used to parameterize it by keeping the generality of flow field intact. Moreover, we implemented, K-means algorithm to initialize the mixture model while Expectation Maximization algorithm is employed for optimization. These mixtures of Gaussian result in the distinct flow patterns (i.e. precisely a sequence of dynamic patterns) for each flow-block. Third, discriminative models such as Conditional Random Field, Hidden Conditional Random Field and Latent-dynamic Conditional Random Field were employed one for each flow-block and were learned from the sequence of dynamic patterns which was then classified for each flow-block as normal and abnormal. Our experiment on our own realistic Data set (Hajj-Umrah DataSet) from crowd in the pilgrimage shows promising results with no scarifying real-time performance for a wide range of practical crowd applications.

Keywords: *Crowd Behavior, Gaussian mixture, K-means technique, conditional Random Field.*

1. INTRODUCTION

In crowded scene, the analysis of behaviors is to inference of joint action which carried out throughout a certain time. Generally talking, there are two levels for analyzing crowd: 1) individual level and 2) global level as depicted in Fig. 1.

At the individual level, the objective is to recognize and understand the human behavior in the crowd scene. Whereas, the goal in the global level is to model the human behavior of any group. A normal behavior (i.e., anomaly) is built by analyzing motion features, characterizing and understanding the behavior in both case. In addition, detecting abnormal behavior represents the action of locating activities, which do not follow the normal behavior. The networks of surveillance video camera are progressively ubiquitous, giving universal data gathering abilities. Every day, some of these cameras stream video of highly congested places such as train stations, airports, subways, and stadium where massive gathering of people is frequently observed. This, together with the growing interest for Wide Area Surveillance (WAS) and Unmanned Aerial Vehicle (UAV) footage, increases the need for effective and robust crowd behavior analysis systems. Most of the time, the main challenge of those systems is to ensure human safety in catastrophic situations [1]. The road map for most surveillance system is shown in Fig. 2 which contains main four stage processes namely object detection, object tracking, object classification and action recognition.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

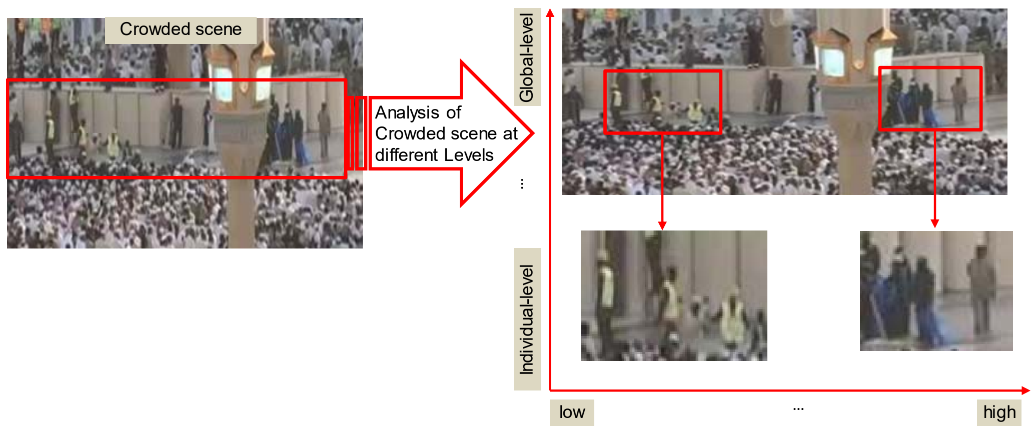


Figure 1: Crowd scene with multiple behaviors in Prophet's Mosque. The source crowded scene showed on the left of the figure, upper right shows the Global-level, and down is the individual-level.

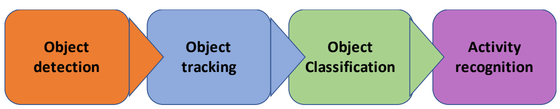


Figure 2: Road Map Automated Surveillance System [1].

Despite the concerted exertions thru using idea researchers, a surveillance assumption about object density is regularly violated from real scene as in Fig.3, which including subjects in various situations. The analysis of crowd behavior has arisen as a challenging study area which should face several non-trivial issues such as massive parallel tracking, occlusion handling, changing illumination, etc. In terms of applicability, crowd behavior analysis and abnormal detection find widespread use in numerous applications like localizing wary activities of entities and aid in alternative circumstances.

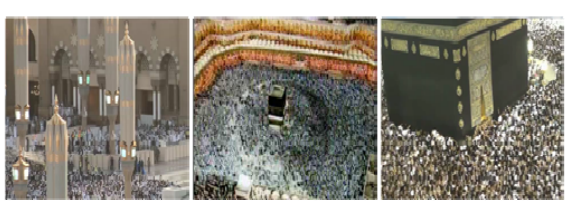


Figure 3: Other Crowded scenes including people in Prophet's Mosque and Makkah Mosque in Saudi Arabia.

In this work, the major contribution is to provide computerized model, where the foreground regions are extracted and segmented video sequence into flow-blocks. Further, K-means clustering is to initialize process and then Expectation Maximization (EM) technique determines maximum likelihood parameters. A distinct representation of flow field is determined by using Gaussian mixtures. In addition, a discriminative model of Conditional Random Fields (CRF), Hidden Conditional Random Field (HCRF) as well as latent-dynamic Conditional Random Field (LDCRF) are employed. More are described in Section 3 (Crowd behavior approach).

Furthermore, the performance for our proposed technique is based on our created own realistic crowd in the pilgrimage (i.e., Hajj and Umrah scenes) dataset. We used six scenarios for crowded behavior classes and Prophet's Mosque in Madinah and Makkah, Saudi Arabia. In each scenario, the crowd density is ranging from light to very crowded. The perambulator direction and locations of walking are selected randomly. Additionally, we based on some passing and fixed normal and abnormal individuals in many scenes to make the scenarios more applicable and realistic. The remains of this work are planned as next: the related work is introduced in Section 2 and Section 3 proposes our crowded behavior analysis approach. Section 4 discusses our experimental results. At end, our technique is concluded as in Section 5.

2. RELATED WORK

From most associated research in crowd analysis field, we institute that crowd behavior analysis and anomaly detection techniques still suffer from limitations and thus could be significantly improved. Among other things, several methods rely on manually-selected thresholds and offer context-specific solutions and thus do not scale to other applications. The area of crowd behavior analysis is roughly spilted into two main categorizes: 1) Behavior analysis with

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

trajectory modeling and 2) behavior analysis with crowd flow modeling.

2.1 Behavior analysis with individual detection

Several techniques and approaches in this category perform scene analysis and anomaly detection based on object tracking. The individual tracks are bundled together to find modes of activities. Among many, an interesting work is reported which tries to track crowds of ants [2] and generate separate trajectories for each object. In contrast, Antonini *et al*. [3] presented the problematic of tracking and detection in crowds by selection discrete models to ordinary behavioral pattern. Additionally, Rodriguez *et al*. [4] presented an approach to track amorphous crowded scene, where numerous crowds behaviors are mapped to different scene locations using Correlated Topic Model (CTM).
Other approaches extract individual moving objects in a crowd to model their individual activities and behavior. One such model is the one by Zhao *et al*. [5] which localize individual pedestrians in crowded scenes using a Bayesian framework. This approach performs well on loose crowds but drops in performance when dealing with highly crowded scenes. Brostow *et al*. [6] presented an unsupervised Bayesian cluster framework to group the trajectory, which represents independent moving entities relied on their proximity of space time. Here again, successful results were stated to low density scenes only. Lately, Stalder *et al*. [7] presented automatic grid-based classifier to detect pattern in crowd relied on local setting where the results are shown on crowds carrying normal walk.

2.2 Behavior analysis with crowd flow

Object extraction and tracking of many moving objects is a fundamentally challenging task. Therefore, many techniques have been stated according to others solution to alleviate tracking and object localization. These solutions rather use low level of gesticulation features like optic flow, gradient, and bits of activity to characterize the crowd dynamics. Early tries, Boghossian *et al*. [8] presented a method to model dynamics scene to give assistance using online illustration and detection of crowd-related emergencies in large crowds. Using optic flow, Andrade *et al.* [9] proposed generative model as fully connected Hidden Markov Model (HMM)) to global level of normal motion objects. The results are demonstrated on synthetic simulation and scenes are taken from top view by detecting outliers as anomaly. With a distinct perspective and for dense crowds, Kratz *et al*. [10] demonstrated the spatio-temporal gradient statistics of cuboid with coupled HMM. Furthermore, Benabbas *et al.* [11] construct probabilistic online model for both orientation and density of flow pattern to learn activity of crowd. In similar, Mahadevan *et al.* [12] presented an approach to model normal dynamic of crowd via the mixtures of textures hypothesis dynamic in which the normal saliency model indicates joint dynamics representation and appearance. Let us also mention the work by Ali and Shah [13] which models crowded scenes by a fluid following Lagrangian Particle Dynamics. The experiments are demonstrated on novel dataset, which contained various anomalies definition as walk in wrong route, automobiles on walking range, etc.

Mehran *et al*. explored crowd behavior via social force models [14], without detection for humane or involving the process tracking [15]. Also, they restrained an interaction force vial estimating the difference between the desired and actual velocity that computed by particle advection to optic flow fields [16]. In addition, they used Latent Dirichlet allocation to estimate the normal behavior distribution relied on social force. Cui *et al.* presented the social action behavior by employing the energy potential interaction [17]. They used the interested points of space time to detect and then track an object by KLT feature tracker to attain humane motion throughout image sequences [18], [19]. The energy probable interaction is computed using the velocity of interest point for space time to clarify whether they will see momentarily [20]. In the interim, further study groups have focused extra on detection for local unusual activities. Mancas *et al*. proposed rarity global quantifiably to choose the irrelevant motion from spatial context via saliency bottom up [21]. Additionally, they restrained an index of saliency to multiple channels that include various speeds and various directions. Finally, the local unusual activities were detected via the map of corresponding saliency. Ihaddadene *et al*. noticed that there are variations in motion for the set of point of interest [22]. Also, they construct a map for heat motion relied on motion intensities. Then, they made the comparison with respect to the local motion variations. Mahadevan *et al.* demonstrated normal behaviors dynamic as well as the appearance information to crowded scene in conjunction with the mixture of dynamic textures [23]. So, they employed saliency spatio-temporal to localize and detect the abnormal actions in crowded videos scene. At end, many attempted has been found to analyze the crowd behavior via extracting the local spatial temporal cuboids with

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

respect to optic flow or features of gradient patterns. Kratz *et al*. investigated enormously crowded video sequences volume via building motion patterns distribution, that apprehended the local spatio-temporal to motion pattern [24]. They modeled the motion pattern to HMM based on distribution [25]. Wang *et al*. restrained variation of intensities frequency related to spatial temporal cuboid time by wavelet transform [26]. They demonstrated that the abnormal area illustrates high frequency through certain period. Furthermore, an optimal spatial temporal cuboid choosing has been addressed [27].

The location and size of cuboid that represent crucial factor effects on feature quality since the spatial temporal has been extracted in small region image frame. Thus, the value of spatial temporal cuboid is enhanced via selecting the local maximum points with respect to Gaussian distribution. Even though the preceding approaches verified their effectiveness in their private experiments, they frequently absorbed solely on global or local abnormal action detection. Besides, we claim that the joint consideration of object size, the flow motion and interactions between patterns can quantifiably signify humane actions in crowded scene and hence they improve the performance of abnormal action detection.

Dong-Gyu Lee *et. al*. projected a new technique to attack these difficulties efficiently throughout unified framework [28]. They first projected the map of motion influence to effectively represent the actions through crowded scene. They supplementary develop a method to detect and localize either global or local for abnormal actions during unified framework using the map of motion influence. D. Zhang *et. al*. [29] projected a method to estimate crowd scene abnormality. This method was relied on Hidden Conditional Random Fields Model (HCRF), in which it rectifies numerous motion objects with respect to direction distribution. They used two various scenarios experimentally to detect the abnormal action motion effectively. The drawback here is to not contain extra motion characteristics like acceleration and velocity for the abnormality detection in crowd scene. V. J. Kok *et. al.* [30] thru a review which affords the readers with evaluation to state-of-the-art approaches in analyzing crowd behavior using biologically as well as physics enthused viewpoints. They gave comprehensive discussions and insights to wider thoughtful of fundamental view of unification biology and physics studies. Rabiee *et al.* [31 presented a new crowd database for crowd behavior and emotion annotations. They trust this database not only can be used as benchmark, but also can open accesses in computer vision community to understand the correlation between two tasks, “crowd behavior understanding’’ as well as “crowd emotion recognition’’. They represented human behaviors by a set of intermediate concepts called emotion attributes which are either manually specified or learnt from training data. So, they proposed a unified framework wherein the emotion attributes can be effectively chosen in a discriminative fashion. But, they do not perform the experiments with some large crowd data sets to validate the proposed methodology in more effective manner.

3. CROWD BEHAVIOR APPROACH

The major contribution of our paper is to incubate computer vision model for the behavior analysis of crowd, where the foreground regions is extracted and segmented video sequence into flow-blocks (Fig. 4). Further, K-means clustering is used to initialize the process and then Expectation Maximization (EM) technique is to determine the maximum likelihood parameters. The second major contribution is to exploit the Gaussian mixtures to determine a distinct representation of flow field which are proceeded as patterns sequence for each flow-block (i.e. for naturalness as patterns or dynamic patterns, interchangeably).

A discriminative model of Conditional Random Field (CRFs), Hidden Conditional Random Field (HCRFs) and latent-dynamic Conditional Random Fields (LDCRFs) with various window size is constructed and assigned as; one state/label to model the sequence of dynamic patterns for each flow-block and to parameterize the crowd behavior at the global and specific level in complex environment. Besides, the overall scene anomaly is computed by finding the statistical ratio between total flow-blocks and the abnormal detected flow blocks. The principle of studying crowd behavior system is to perform behavior understanding and anomaly detection by analyzing motion features.

Journal of Theoretical and Applied Information Technology
15th June 2018. Vol.96. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

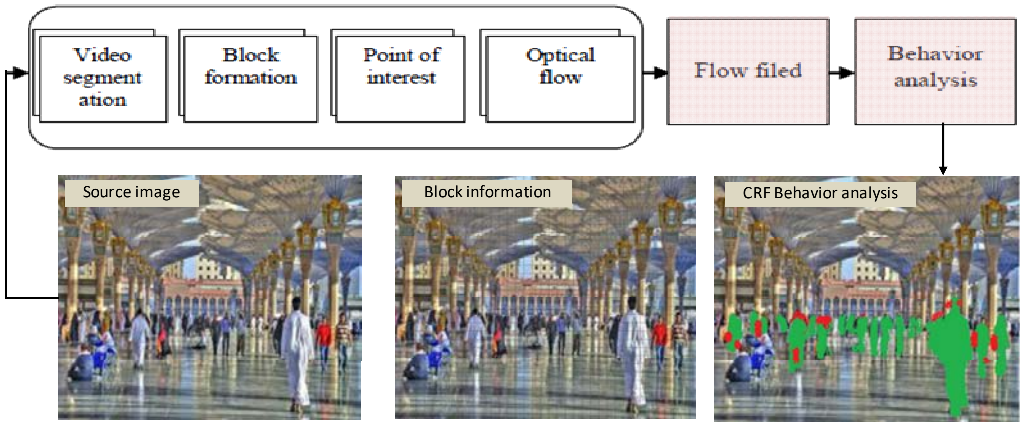


Figure 4: Outline of the crowd behavior approach.

In general, the analysis of crowd behavior is divided in normal behavior and abnormal behavior. Broadly speaking, there are two levels of crowd analysis; individual level and global level. At the individual level, the goal is to extract and understand behavior of each moving object in the crowd. At the global level, the goal is to model the behavior of the group. In both cases, one can perform behavior understanding and anomaly detection by analyzing motion features and characterizing so-called" normal behavior". In contrast, detecting" anomaly" or" abnormal behavior" refers to the action of locating activities, which not imitate to" normal behavior" or drop in their own labeled class.
Fig. 5 shows how to detect and extract the object features from the motion fields and then clustering it by k-mean algorithm. Therefore, the constructing feature vector is assign to CRFs and other their types classifier for behavior analysis where the red color refers to abnormal while green color is to normal behavior.

3.1 Segmentation and Block Creation

Background subtraction technique is an openly used approach to detect the unusual motion in a scene, which involves comparing each new frame to a designed model against the scene background. It is worth mentioning that, Gaussian Mixture Models (GMM) are an example of a larger brand of density models that have many functions as collective components [32]. Formally speaking, Let X_t represents a pixel at existing frame, as well as K is to the distributions number. Thus, each pixel can be classified separately by a Gaussian mixture K as follows;

$$p(X_t) = \sum_{i=1}^{K} \omega_{i,t} \cdot \eta(X_t; \mu_{i,t}; \Sigma_{i,t}) \qquad (1)$$

where η refers to a Gaussian probability density function. $\mu_{i,t}, \Sigma_{i,t}$ and $\omega_{i,t}$ represent the mean, covariance and prior probability estimation (weighting function) of i^{th} component, respectively.

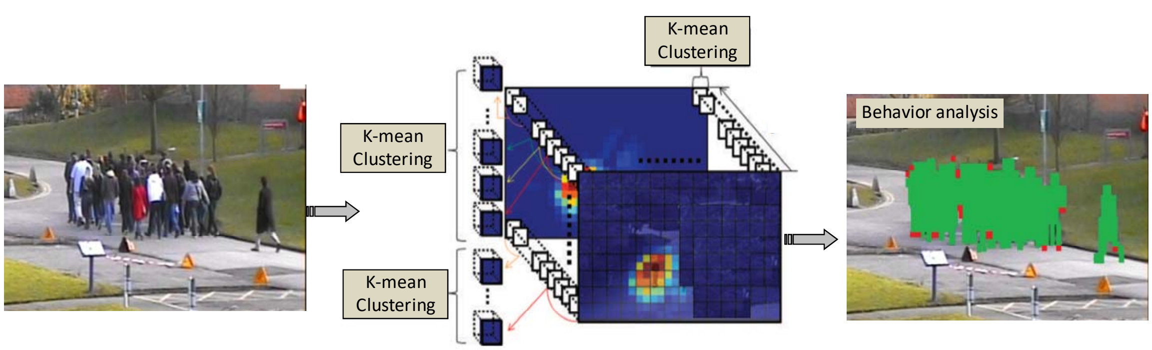


Figure 5: Principle of crowded behavior system.

A constructive procedure is used to regulate the numbers of component inevitably based on a principle of maximizing likelihood [32].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Furthermore, the background procedure is updated and then optimized relied on the error function minimization E as follow;

$$E = -\sum_{n=1}^{N} ln\Big(\sum_{i=1}^{K} \eta(X;\mu_i;\Sigma_i)\cdot\omega_i\Big) \quad (2)$$

where N represents the data points number X_n. here, the background distributions still on top related to the smallest variance with threshold γ = 0.5. At end, all pixels X, that do not match any component are greatest runners to be marked as foreground. The reader is to reference [33] for details. We employ Expectation Maximization (EM) procedure as unusual case of Maximum Likelihood (ML) technique to estimate the mixture model parameters that turns finest to the dataset over ML Video sense [34]. Based-magnitude optic flow pruning contains two phases for separating relevant from irrelevant flow vector [35]. In First phase, all vectors with magnitudes that either very small or very large relatively are removed [36].

So, we used two predefined maximum and minimum thresholds to rectify the flow vectors filtering. Momentarily talking, given two thresholds ρ_1 and ρ_2, the flow vector $\vec{v}_{op} = [x, y]^T$ is first recognized as true when satisfies the validity limitation: $\rho_1 < \|\vec{v}_{op}\| < \rho_2$ such that $\|\cdot\|$ represents the magnitude of flow vector according Euclidean metric; else, it is supposed to refer the raucous flow component and hence it is detached. In the second phase, the vector is considered as true component when the Euclidean between the object flow center and the vector that be analyzed with no a definite threshold τ doesn't outdo. Mathematically, it is denoted as next;

$$\|\vec{v}_{op} - \vec{z}\| < \tau \quad (3)$$

where $\vec{z}$ represents motion area centroid. for our experiment, the set-up values of ρ_1 = 4; ρ_2, = 22 and τ at 24% of the average width w of image and the average height h; $\ell = (w+h)/2$ give an overall good pruning performance. Thus, the optical flow between the sequential frames of image sequences is computed. Figure 6 show the optical flow estimation for a crowd scene in addition to segment the region of interest.

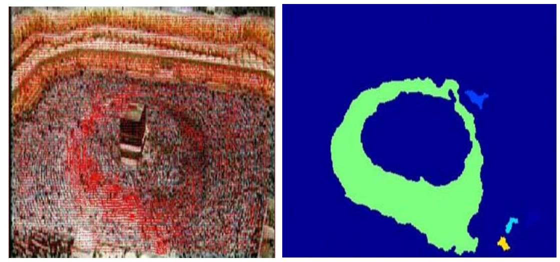


Figure 6: The observed flow field with its respective magnitude as well as the segmentation of the region of interest.

3.2 Parameterizing block flow field

K-means cluster technique is active to initialize and calculate the parameters model to every cluster. The impetus behindhand using cluster mechanism like k-means is to state the features from each block flow field. From the sequence of motion influence map (Fig. 7), we concatenate the spatio-temporal motion influence vectors extraction to further analysis the motion behavior using k-means clustering techniques. Moreover, we used k-means function for many pros that brand it extra widespread in terms of use like convergence speed, adaptability, convergence speed and scalability for sparse data. The core idea of using k-means function, which based on Euclidean distance in-between clusters centroid as well as all points of their clusters is very modest. As, every point is firstly allocated to one initialized cluster.

After that, the centroid of each cluster is re-estimated using the mean point of capable points of cluster. Thereby, this process is repeated till achieving convergence.

3.3 Classification: Conditional Random Field

The discriminative Conditional Random Field (CRF) is an undirected graph model, which are sophisticated to label sequential data [37]. For certain emission, there is trade-off to the weight in every function of feature, because the discriminative model of CRF uses only exponential distribution to model all reference labels [38]. Correctly talking, for certain emission sequence, the label sequence probability y is estimated as nest;

$$p(y|x,\theta) = \frac{1}{Z(x,\theta)} \exp\Big(\sum_{i=1}^{n} F_\theta(y_{i-1}, y_i, x, i)\Big) \quad (4)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

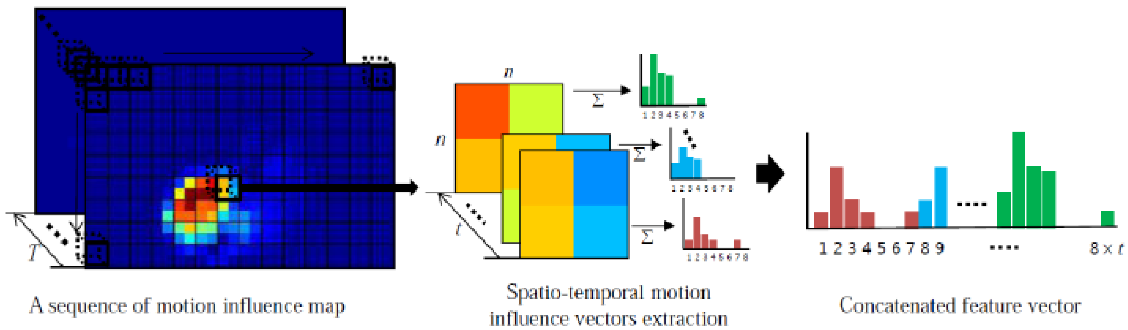


Figure 7: Spatio-temporal feature vector extraction map.

Such that; the parameter θ is equal $(\lambda_1, \lambda_2, \dots, \lambda_{N_f}; \mu_1, \mu_2, \dots, \mu_{N_g})$. N_f is to the number of transition function of features, *Ng* denotes to the number of label function of features as well as *n* represents an observation sequence length *x*. Thus, F_θ is treated as follows;

$$F_\theta(y_{i-1}, y_i, x, i) = \sum_f \lambda_f t_f(y_{i-1}, y_i, x, i) + \sum_g \mu_g s_g(y_i, x, i) \quad (5)$$

Such that $t_f(y_{i-1}, y_i, x, i)$ represents transition function of feature to location *i-1* and *i*. $s_g(y_i, x, i)$ refers to the labels function of feature to location *i*. λ_f as well as μ_g represent the transition weights and labels function of features, respectively. $Z(x, \theta)$ is to a normalized factor in which it is estimated as next;

$$Z(x, \theta) = \sum_y \exp\left(\sum_{i=1}^{n} F_\theta(y_{i-1}, y_i, x, i)\right) \quad (6)$$

Other classifiers such as Latent-dynamic Conditional Random Field (LDCRF) and Hidden Conditional Random Field (HCRF) enforce an invariant process according to their hidden states. The main motivation behind using CRFs is to treat the weakness of the model of directed graph. But contrary there has a bias problem as well as they have not a capability to train and state an internal sub-structure of meaningful pattern sequence.

So, CRFs is extended to provide other techniques so-called HCRFs and LDCRFs which automatically model the internal sub-structure of a specific pattern [39], [40], [41], [41]. The pro of HCRF is to prototypical a local substructure among states based on hidden variable. But, it has no the ability to perform the dynamic process among their states. Furthermore, LDCRFs treats this problem by modeling both local substructure and dynamic training processes among their states in addition to classify the unsegment meaningful pattern. In short, we can say that the LDCRFs combine the advantages of both CRFs and HCRFs.

4. EXPERIMENTS DISCUSSION

4.1 Hajj-Umrah DataSet

We created our own realistic crowd in the pilgrimage (i.e., Hajj and Umrah scenes) dataset. The data collected from fixed cameras stated in different crowd places, this dataset contains 120 video clips (i.e., nearly 14400 individual frames) with 640×480 pixels' image resolution at 30 FPS. The duration of each snapshot is approximately 4 sec. In our dataset, we used six scenarios for crowded behavior classes in Makkah and Prophet's Mosque in Saudi Arabia (Fig. 8).

Figure 8: The set of scenes collected from various crowd places in the Makkah Mosque in (Masjid al-Ḥarām) as well as the Prophet's Mosque in Medina.

In each scenario, the crowd densities are ranging from light to identical crowded. The perambulator direction and locations of walking are selected randomly. Additionally, we based on some passing and fixed normal and abnormal individuals in

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

many scenes to make the scenarios more applicable and realistic. Fig. 8 explores example frames from Hajj and Umrah scenes of dissimilar categories. We divided own dataset into one third to test set and two thirds learning set.

4.2 Experimental results

In our proposed research, the optic flow analysis depended on a motion-related descriptor is used. Here the computations of optical flow are sensitive the noise of background and vary in scale as well as the direction of motion. Additionally, we consider the moving pixels subject to the time. For these restrictions, the optical flow value to be expected less unsuitable or suitable features for the analysis of motion. To alleviate these difficulties, the distribution characters of flow are used as a feature to describe the motion. It is a truth; the motion activity for an individual in the static background scene can be fully characterized using the profile of own self-induced optical flow.

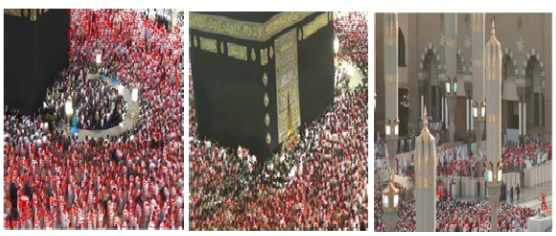


Figure 9: Three different frame samples from different scene showing the observed flow field with its respective magnitude.

In Fig. 9, optic flow samples to patterns in own dataset viewing the observed flow field with its respective magnitude.

In Table 1, a qualitative representation of CRFs, HCRFs and LDCRFs for our own Hajj and Umrah dataset was provided which indicates the total number of detected blocks and respective Normal (NB) and abnormal (AbNB) detected behaviors in every time instance.

The probability of abnormal and normal behavior analysis is estimated using confusion matrix to every label in dynamics crowd scene (Table. 1). It is being noted that, the diagonal elements of this confusion matrix represent the percentage probability to every label in group. Moreover, the misclassification among classes were displayed using non-diagonal foundations that were noticed respective to memorable motion field at the parts of pattern legs as compared to head and the body of the object's subjects.

Here, the learning of CRF, HCRF and LDCRF techniques used the criteria of gradient ascent with respect to Broyden-Fletcher-Goldfarb-Shanno (BFGS). Additionally, window size ranging from 1 to 8 with 200 iterations is to perform the learning processes till parameters convergence for CRF, HCRF and LDCRF. The window size of one represents vector of size three; the current frame, one earlier frame and one coming frame. But, an inferencing (i.e., testing) process is round as fast of three models in ordered of seconds to each scene sequences.

We have three errors types, which so-called deletion, insertion and substitution. The error of deletion will happen if the classifier fails to detect the specific behavior. In insertion error, the classifier detects nonexistent value because the probability of observation features of the current frame is equal to zero for given behavior features. The substitution error will occur when the classifier classifies the action falsely (i.e, classify normal behavior as abnormal behavior and vice versa). This error dues to the extracted features are assigned to other behavior features.

To estimate the ratio of recognition, we don't consider the insertion error entirely (Eq. 7). Nevertheless, insertion error is a strong decision to determine the action behaviors rather than deletion and substitution error.

$$Recognition\ ratio = \frac{\#\ recognized\ actions}{\#\ test\ actions} \times 100 \quad (7)$$

The error of deletion affects on the recognition ratio directly while the error of insertion does not. Furthermore, we use another measurement that called reliability to illustrate the outcome of insertion error as in the next equation;

$$Reliability = \frac{\#\ correctly\ recognized\ actions}{\#\ test\ actions + \#\ Inseration\ errors} \times 100 \quad (8)$$

Table 1: Confusion matrix of Hajj and Umrah behavior results.

Event	CRFs		HCRFs		LDCRFs	
	NB	AbNB	NB	AbNB	NB	AbNB
NB	95.2	4.8	96	4	97.1	2.9
AbNB	5.1	94.9	4.3	95.7	3.8	96.2

Journal of Theoretical and Applied Information Technology
15th June 2018. Vol.96. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

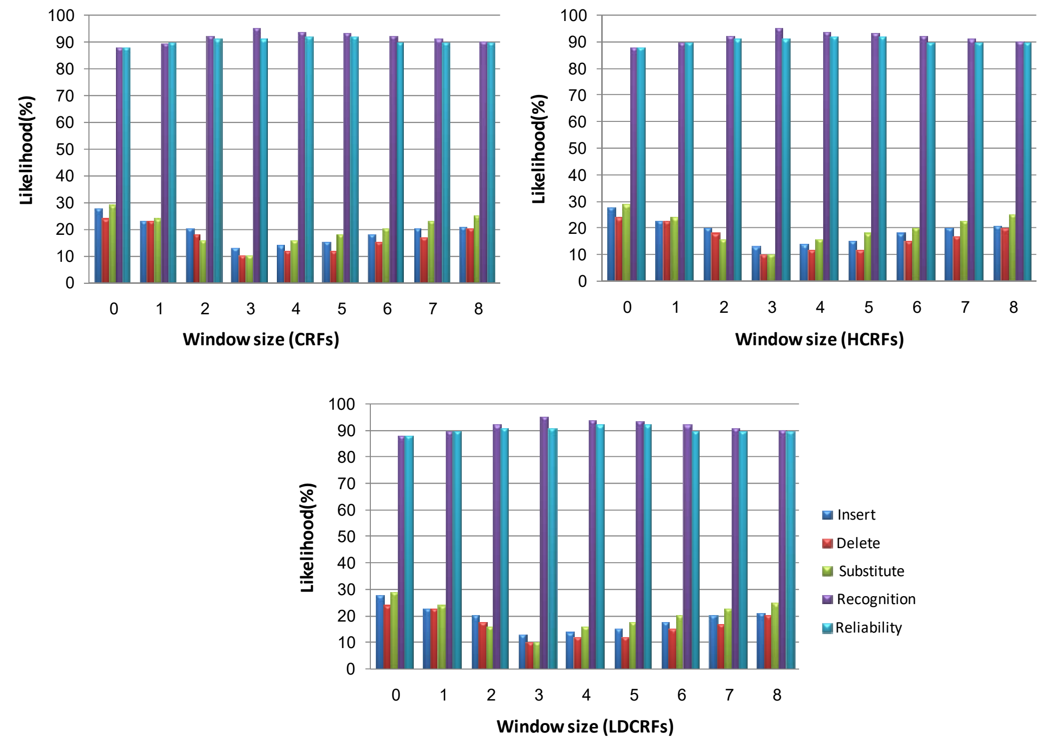


Figure 10: Hajj and Umrah dataset recognition rate, reliability and three types of errors for CRF, HCRF in addition to LDCRF with respect to various ranging of window size from 0 to 8.

Consequently, the projected framework achieved well recognition at window size 3, in which it was recognizing tested human behavior automatically with 95.2%, 96%, 97.1% fto CRF, HCRF and LDCRF, correspondingly (Table 1).

It was noted that, LDCRF is higher than HCRF and CRF related to tested dataset (Fig. 10). Besides, the recognition rate of behavior analysis is enhanced firstly as window size increases nevertheless reduces as window size supplementary increase (Fig.10).

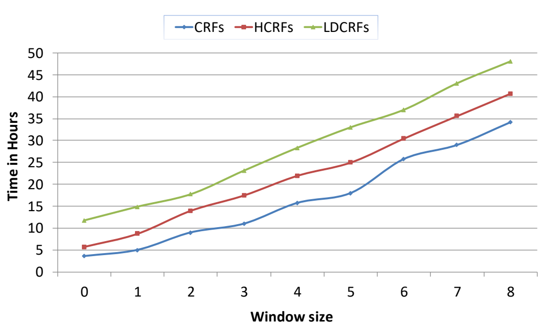


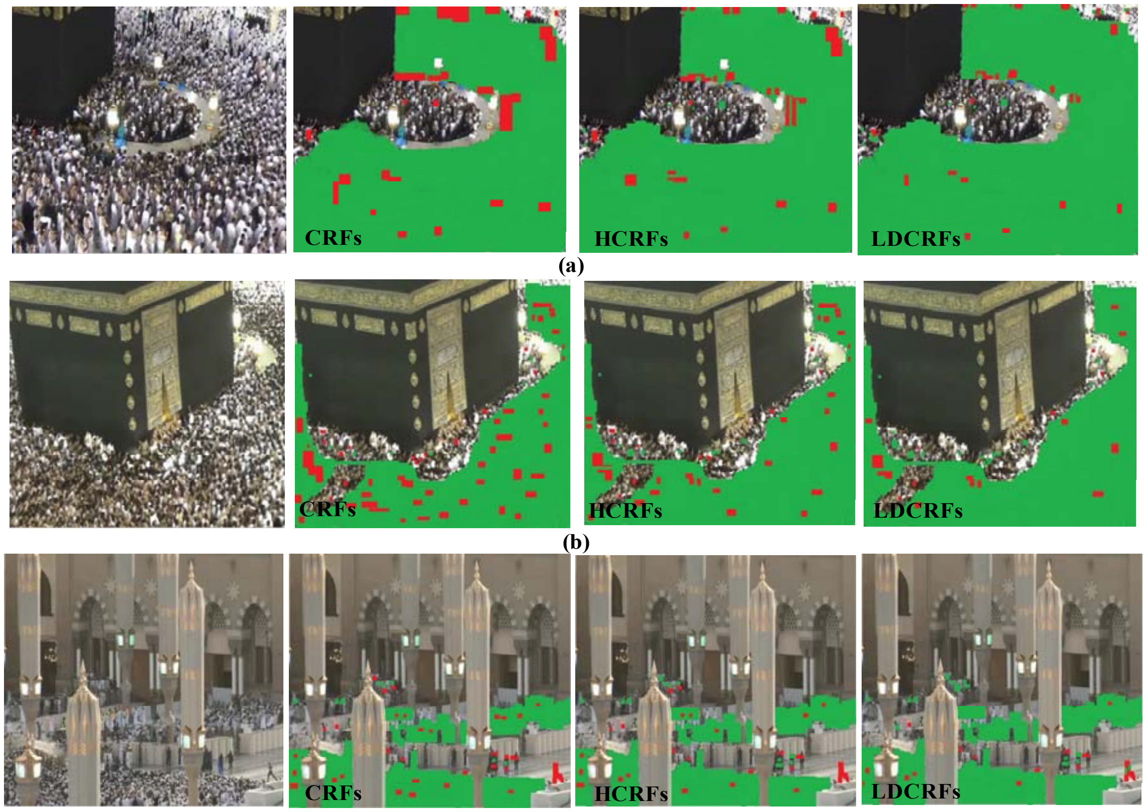
Figure 12: Effecting cost in terms of time for the training processes of Hajj and Umrah dataset subject to LDCRF, HCRF and CRF with various ranging window size from 0 to 8

The crowded behavior results on Hajj and Umrah dataset is outcome in Fig. 11, in which the detection of normal behavior is assigned using green spots and abnormal behavior marked with red spots. The time cost of CRF, HCRF and LDCRF models on Hajj and Umrah dataset over various window size is proportional summarized and illustrated in Fig. 12. As well, it illustrated that the model of LDCRF permit to represent hidden states, which are usually more expensive than their standard counterparts.

4.3 Evaluation criteria

We have explored the criteria of anomaly detection relied on the classification results. To evaluate our system, the classification outcomes will be analyzed using quantitative analysis. In addition, we put on Social Entropy (SE) to carry out the problems of optic flow [41]. The main motivation behind using SE is to empirically find the quantitative metric as well as permits us to run the optic flow noise. To detect anomaly, we suppose that the normal and abnormal labels is somehow fuzzy for a scene. Moreover, we have given behaviors (i.e., normal and abnormal) to scene analyst to take aid from visual representation.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



(c)

Figure 11: Crowded behavior results on Hajj and Umrah dataset in which the detection of normal behavior is specified using green spots and an abnormal behavior is indicated using red spots. (a) Source frame close-up perspective of the Kaaba with its results in CRFs, HCRFs and LDCRFs respectively. (b) Source frame Far from the Kaaba view. (c) Entry view of individuals to the Prophet's Mosque through the peace door with its results in CRFs, HCRFs and LDCRFs respectively.

Table 2: Crowd behavior recognition using CRFs, HCRFs and LDCRFs

Datasets	CRFS				HCRFs				LDCRFs			
	NB		AbNB		NB		AbNB		NB		AbNB	
	pre.	rec.	pre.	rec.	pre.	rec.	pre.	rec.	pre.	rec.	pre.	rec.
Hajj-Umrah	95	95.2	94.1	94.9	95.1	96	94.4	95.7	96.3	97.1	95.6	96.2

Also, we have provided a selection of defining a threshold to sake the automization as in Eq. 9. It is being noted that the abnormal behavior will be above 50%.

$$Anomaly = \frac{AbNB}{AbNB + NB} * 100 \tag{9}$$

where *Anomaly* is to the total abnormal for scene behavior. It is calculated as a ratio to the total detected flow-blocks like abnormal comparative to the total of abnormal and normal that classified flow-locks. Thus, we have an ability to identify and analyze the status of the crowded scene behavior at each time instance. Our system is evaluated based on the corrected behavior correspond to every flow-block in the sense. For such evaluation, we firstly created a ground truth for analysis the normal as well as abnormal flow range on learning dataset. This is run using a training procedure to label behavior either normal or abnormal for every block since it is relatively monotonous to manually allocate labels on every flow-block. Thus, the performance of system is measured using the precision (*pre.*) and recall (*rec.*) calculating as in Eq. 10 and Eq. 11.

$$pre. = \frac{number\ of\ corrected\ behaviors}{number\ of\ established\ behaviors} \tag{10}$$

$$rec. = \frac{number\ of\ corrected\ behaviors}{number\ of\ actual\ behaviors} \quad (11)$$

where actual behaviors refer to the available behaviors (i.e., normal and abnormal) in ground truth. Table 2 illustrates the *precision* and *recall* measurement of abnormal and normal behavior recognition with CRFs, HCRFs and LDCRFs to every behavior class. Every column boons the respective analysis for *precision* as well as *recall* of abnormal and normal classes.

5. CONCLUSION

In our work, we explored the novelty technique and localized the anomalies respective to the crowd behaviors of individual's hasty rakishness. In first, the foreground regions were extracted and segmented video sequence into flow-blocks. K-means clustering was used to adjust process and then Expectation Maximization technique determined maximum likelihood parameters. A discriminative model of CRFs, HCRFs and LDCRFs was constructed and assigned as; one state/label to model dynamic pattern sequences to every flow-block and parameterizing the crowd behavior at global and specific level in complex situation. Besides, the overall scene anomaly was computed by finding the statistical ratio between total flow-blocks and the abnormal detected flow-blocks.

Likewise, the performance of method relied on our own Hajj and Umrah dataset. It is being noted that, the recognition rate of LDCRFs is more robust than CRFs and HCRFs in detecting overall and specific crowd behaviors with no losing real-time performance to varied range of applied crowd applications. The main motivating factor for this work is to apply it in the future (i.e., future recommendation) during the season of Hajj and Umrah, for example, are the restrictions of the ability of technical surveillance systems make pilgrims behavior and provide useful information and sound for safety and security official. This requires an installation network of cameras in Makkah and develops computer vision algorithms for the behavior analysis of crowd during Hajj and Umrah by incorporating the human psychological studies.

ACKNOWLEDGMENT

This work was reinforced via King Abdulaziz City of Science & Technology (KACST), KSA under Grant AC- 35-92. We would like to thank them to provide us an opportunity to perform this research.

REFERENCES

[1] D. Helbing, A. Johansson, and H. Al-Abideen, "The dynamics of crowd disasters: An empirical study", *Phys. Rev*, 75:046109, 2007.

[2] Z. Khan, "MCMC-based particle filtering for tracking a variable number of interacting targets", *IEEE Transaction of Pattern Analysis and Machine Intelligence*, 27(11), 2005, pp. 1805–1918.

[3] G. Antonini, S. Martinez, V. Santiago, M. Bierlaire, and J. Thiran, "Behavioral priors for detection and tracking of pedestrians in video sequences", *International Journal of Computer Vision*, 69(2), 2006, pp.159–180.

[4] M. Rodriguez, S. Ali, and T. Kanade, "Tracking in unstructured crowded scenes", *In ICCV*, 2009, pp. 1389–1396.

[5] T. Zhao and R. Nevatia, "Bayesian human segmentation in crowded situations", *In Proceeding of Computer Vision and Pattern Recognition*, 2003, pp. 406–413.

[6] G. Brostow and R. Cipolla, "Unsupervised Bayesian detection of independent motion in crowds", In *Proceeding of Computer Vision and Pattern Recognition*, 2006, pp. 594–601.

[7] S. Stalder, H. Grabner, and L. Gool, "Exploring context to learn scene specific object detectors", In IEEE *International Workshop on Performance Evaluation of Tracking and Surveillance*, Miami, USA, 2009, pp. 63–70.

[8] A. Boghossian and A. Velastin, "Motion-based machine vision techniques for the management of large crowds", *In Proceedings of 6th IEEE International Conference on ICECS Electronics, Circuits and Systems*, Volume 2, 2002, pp. 961–964.

[9] E. Andrade, B. Scott, and R. Fisher, "Hidden markov models for optical flow analysis in crowds", *In Proceedings of the 18th International Conference on Pattern Recognition*, 2006, pp. 460–463.

[10] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models", *In IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[11] Y. Benabbas, N. Ihaddadene, and C. Djeraba, "Global analysis of motion vectors for event detection in crowd scenes", *In Performance*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Evaluation of Tracking and Surveillance workshop at CVPR, 2009, pp. 109–116.

[12] V. Mahadevan, W.X. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes", *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1975–1981.

[13] S. Ali and M. Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis", *In IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–6.

[14] D. Helbing and P. Molnar, "Social Force Model for Pedestrian Dynamics," *Physical Review E*, Vol. 51, No. 5, 2007, pp. 4282-4286.

[15] R. Mehran, A. Oyama, and M. Shah, "Abnormal Crowd Behavior Detection Using Social Force Model," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Miami, USA, June 20-25, 2009, pp. 935-942.

[16] S. Wu, B. Moore, and M. Shah, "Chaotic Invariants of Lagrangian Particle Trajectories for Anomaly Detection in Crowded Scenes," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, USA, June 13-18, 2010, pp. 2054-2060.

[17] X. Cui, Q. Liu, M. Gao, and D. N. Metaxas, "Abnormal Detection Using Interaction Energy Potentials," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Colorado, USA, June 20-25, 2011, pp. 3161-3167.

[18] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, "Learning Realistic Human Actions from Movies," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, USA, June 23- 28, 2008, pp. 1-8.

[19] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proc. 7th International Joint Conference on Artificial Intelligence*, San Francisco, USA, Aug. 3-9, 1981, pp. 674-679.

[20] S. Pellegrini, A. Ess, K. Schindler, and L. V. Gool, "You'll Never Walk Alone: Modeling Social Behavior for Multi-Target Tracking," *Proc. 12th IEEE International Conference on Computer Vision*, Kyoto, Japan, Sep. 29 - Oct. 2, 2009, pp. 261-268.

[21] M. Mancas, N. Riche, J. Leroy, and B. Gosselin, "Abnormal Motion Selection in Crowds Using Bottom-Up Saliency," *Proc. 18th IEEE International Conference on Image Processing*, Bruxelles, Belgium, Sep. 11-14, 2011, pp. 229-232.

[22] N. Ihaddadene and C. Djeraba, "Real-time Crowd Motion Analysis," *Proc. 19th IEEE International Conference on Pattern Recognition*, Florida, USA, Dec. 8-11, 2008, pp. 1-4.

[23] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly Detection in Crowded Scenes," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, USA, June 13-18, 2010, pp. 1975-1981.

[24] L. Kratz and K. Nishino, "Tracking Pedestrians Using Local Spatio-Temporal Motion Patterns in Extremely Crowded Scenes," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 5, 2012, pp. 987-1002.

[25] L. R. Rabiner," A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE Reading in Speech Recognition*, 1989, pp. 267-296, ISBN: 1-55860-124-4.

[26] B. Wang, M. Ye, X. Li, F. Zhao, and J. Ding, "Abnormal Crowd Behavior Detection Using High-Frequency and Spatio-Temporal Features," *International Journal of Machine Vision and Applications*, Vol. 23, No. 3, 2012, pp. 501-511.

[27] B. Wang, M. Ye, X. Li, and F. Zhao, "Abnormal Crowd Behavior Detection Using Size-Adapted Spatio-Temporal Features," *International Journal of Control, Automation, and Systems*, Vol. 9, No. 5, 2011, pp. 905-912.

[28] D.-. Lee, H.-I Suk, S.-K. Park, and S.-W. Lee," Motion Influence Map for nusual Human Activity Detection and Localization in Crowded Scenes," *IEEE Transactions on Circuits and Systems for Video Technology*, 2015, pp. 1-12.

[29] D. Zhang, K. Xu, Y. Lu, C. Pan and H.Peng, " Abnormal Crowd Motion Detection with Hidden Conditional Random Fields Model," *International Journal of Multimedia and Ubiquitous Engineering*, Vol. 10, No. 10, 2015, pp. 91-98.

[30] V. J. Kok, M. K. Lim, and C. S. Chan," Crowd Behavior Analysis: A Review where Physics meets Biology", *Neurocomputing 177*, 2016, pp. 342-62.

[31] H. Rabiee, J. Haddadnia and H. Mousavi," Crowd behavior representation: an attribute-based approach", *SpringerPlus*, 5(1):1179, 2016.

[32] S. J. McKenna, Y. Raja, S. Gong, "Tracking colour objects using adaptive mixture models," *Journal of Image and Vision Computing*, Vol. 17, 1999, pp. 225–231.

[33] M. Elmezain, Invariant color features–based foreground segmentation for human-computer interaction, *Math Meth Appl Sci.* 2017*;1–10.*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[34] M. H. Yang and N. Ahuja, "Gaussian Mixture Model of Human Skin Color and Its Applications in Image and Video Databases", *In SPIE/EI&T Storage and Retrieval for Image and Video Databases*, 1999, pp. 458-466.
[35] A. Fathi, G. Mori, "Tracking Colour Objects Using Adaptive Mixture Models", *IEEE Conf. on Computer Vision and Pattern Recognition*, 2008, pp. 1-8.
[36] K. Liu, Q. Du, H. Yang, B. Ma, "Optical Flow and Principal Component Analysis-Based Motion Detection in Outdoor Videos", *EURASIP Journal on Advances in Signal Processing*, 2010, pp. 1-6.
[37] J. Lafferty, A. McCallum and F. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data", *Conf. on ICML*, 2001, pp. 282-289.
[38] H. Yang, S. Sclaroff and S. Lee, "Sign Language Spotting with a Threshold Model Based on Conditional Random Fields", *IEEE Trans. on PAMI*, Vol. 31(7), 2009, pp. 1264-1277.
[39] A. Gunawardana, M. Mahajan, A. Acero, J. C. Platt, "Hidden Conditional Random Fields for Phone Classification", *Proceeding of European Conf. on Speech Communication and technology*, 2005, pp. 1117-1120.
[40] A. Quattoni, S. Wang, L. P. Morency, M. Collins, T. Darrell, "Hidden Conditional Random Fields", *IEEE Trans. on PAMI*, Vol. 29, No. 10, 2007, pp. 1848-1852.
[41] A. Lerner, Y. Chrysanthou, and D. Lischinski, "Crowds by Example," *Computer Graphics Forum*, Vol. 26, No. 3, 2007, pp. 655-664.
[42] S. El-etriby, M. Miraoui, and M. Elmezain "Crowd Behavior Analysis Using Discriminative Models in Video Sequences", *International Journal of Computational and Theoretical Nanoscience*, Vol. 14, No. 6, pp. 2017, pp. 2706–2714.