T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

ABNORMAL BEHAVIOR DETECTION IN SURVEILLANCE SYSTEMS

AYBARS TOKTA
A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE
DEPARTMENT OF ELECTRONIC ENGINEERING

GEBZE
2016

T.R.

GEBZE TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

ABNORMAL BEHAVIOR DETECTION IN SURVEILLANCE SYSTEMS

AYBARS TOKTA

A THESIS SUBMITTED FOR THE DEGREE OF

MASTER OF SCIENCE

DEPARTMENT OF ELECTRONIC ENGINEERING

THESIS SUPERVISOR

ASSIST. PROF. DR. ALİ KÖKSAL HOCAOĞLU

GEBZE

2016

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

GÜVENLİK SİSTEMLERİ ARACILIĞIYLA KALABALIKLARDA ANORMAL DURUM TESPİTİ

AYBARS TOKTA
YÜKSEK LİSANS TEZİ
ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

DANIŞMANI
YRD. DOÇ. DR. ALİ KÖKSAL HOCAOĞLU

GEBZE
2016

GEBZE TEKNİK ÜNİVERSİTESİ	**YÜKSEK LİSANS JÜRİ ONAY FORMU**

GTÜ Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 15/06/2016 tarih ve 2016/37 sayılı kararıyla oluşturulan jüri tarafından 24/08/2016 tarihinde tez savunma sınavı yapılan Aybars TOKTA'ın tez çalışması Elektronik Mühendisliği Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE
(TEZ DANIŞMANI) :Yrd. Doç. Dr. Ali Köksal HOCAOĞLU

ÜYE :Doç. Dr. Koray KAYABOL

ÜYE :Yrd. Doç. Dr. Ulaş VURAL

ONAY

Gebze Teknik Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun

…..…/…..…/..…… tarih ve ……/…….. sayılı kararı.

İMZA/MÜHÜR

ÖZET

Globalleşen dünyada sürekli artan insan nüfusu ve önemli yerlerin sayısı, akıllı güvenlik sistemlerine olan ihtiyacı arttırmaktadır. Bu sistemler operatörlere büyük yardım sağlamalarının yanında ters bir durum olması durumunda olaya müdahale süresini kısaltıp can ve mal kayıplarını azaltabilecektir. Bu çalışmada kalabalıkların anormal davranışlarının otomatik tespiti üzerinde çalışılmıştır. Literatürde bu problemin çözümünde geleneksel optik akış temelli yöntemler kullanılmaktadır. Biz ise eğme temelli optik akış ve etki haritası yaklaşımını kullanarak bir algoritma geliştirdik.

Anahtar Kelimeler: Anormal Durum Tespiti, Optik Akış, Sahne Normalizasyonu, Etki Haritası.

SUMMARY

Public safety has become an important issue in recent years. Developing smart systems to detect abnormal crowd behavior is crucial to take control of the situation as soon as possible. There have been many studies related to topic. Most of these rely on traditional optical flow algorithms. In this study, we propose a novel algorithm based on High Performance Optical Flow and Influence Map. We validate our algorithm in publicly known dataset and compare the detection performance of our method with some other well known methods in the literature.

Key Words: Abnormal Human Behavior Detection, Optical Flow, Scene Normalization, Influence Map.

ACKNOWLEDGEMENTS

This thesis is not possible without the support, guidance and love from so many people around me.

I would like to thank to my parents for their love and support during my whole life. I also thank to my mentor Asst. Prof. Köksal Hocaoğlu for helping me whenever I struggle and couraging me to be better researcher. Finally, I would like to thank to my colleague Assist. Researcher Hasan Huseyin Sönmez for being a perfect friend and helping me to gain time in developing my method.

TABLE of CONTENTS

LIST of ABBREVIATIONS and ACRONYMS

Abbreviations and Acronyms		**Explanations**
μ	:	Mean
σ	:	Variance
Σ	:	Covariance
∇	:	Gradient
τ	:	Relaxation Parameter
v_i	:	Velocity
A	:	Grid
AF	:	Abnormal Frames
DM	:	Distance Matrix
F	:	Force
GAE	:	Global Abnormal Event
GMM	:	Gaussian Mixture Model
GTU	:	Gebze Technical University
Hist	:	Histogram
KLT	:	Lucas Kanade Tomasi
LAE	:	Local Abnormal Event
m	:	Mass
N	:	Normal Distribution
NF	:	Normal Frames
$R_{influence}$	:	Grid Influence Range Parameter
$R_{feature}$	:	Grid Temporal Range Parameter
sgn	:	Signum
std	:	Standard Deviation
SEV	:	Scene Energy Value
SSD	:	Sum of Squared Distances
u	:	Displacement along the x-axis
UMN	:	University of Minnesota
v	:	Displacement along the y-axis

LIST of FIGURES

LIST of TABLES

1. INTRODUCTION

In the 21st century, the number of surveillance systems has increased enormously in order to ensure public security. Just like in many areas, automation has become inevitable in security systems due to increased number of population as well as the number of the surveillance cameras. Especially in dangerous circumstances, it is crucial to intervene to situation as soon as possible. Recent years, developing intelligent security systems has been an appealing area among computer vision researches. Some have tried to understand individual's activities whereas the others approach the crowd as a whole entity. Recent works have showed that detecting abnormal behavior is one of the key goals when it comes to develop an intelligent security system.

1.1. Purpose and Content of The Thesis

In this work, we propose a novel method to detect abnormal human behavior in crowded areas. This algorithm can identify whether or not crowd expose unusual behaviors such as running or escaping from a danger. Doing that, it can produce an alert when an abnormal situation occurs which could help reducing the number of casualties.

The problem mainly involves human motion that can change abruptly during an extraordinary situation. Thus, optical flow is an important parameter in detecting the abnormal events in video frames. So far, the previous works utilized traditional optical flow methods whose performance is vulnerable in many real time applications. There have been many proposed optical flow algorithms in the literature which we found that their overall performance is degraded mainly due to the optical flow algorithm that they utilized. Because of that, our method utilizes state of art coarse to fine optical flow information [1] as an input our detection algorithm.

To detect and localize the abnormality in video frames we create a novel Influence map method which uses optical flow vectors amplitude and phase information to produce an energy matrix of two consecutive video frames. Commonly used dataset as well as our own dataset are used to calculate the overall performance of the proposed method.

1.2. Motivation

Increasing population of countries and technology bring globalization as well as serious security demand. For the sake of this demand, understanding the human behavior from video frames has always been appealing and challenging topic among computer vision researchers. In surveillance cameras, most of the time the number of people is so high that it is hard to track each person individually. We aim to design an effective method to find frames where people are in panic, using holistic approach rather than focusing each individual. Accomplishing that would enable to create smart security systems that can understand crowd activity so that it can produce an alert as soon as something odd happens in the scene, which could then prevent greater undesired situations.

1.3. Thesis Organization

This thesis is organized in five parts. In the second section, we present the works related to abnormal activity detection problem as well as optical flow methods, a key factor in understanding the motion information. We present our proposed method in the third section of this thesis where we mention the algorithm steps thoroughly. The fourth part is the test and analysis part where we evaluate our methods performance on various datasets and make comparison between our methods and the state of art methods in the literature.

2. LITERATURE SUMMARY

Providing a better security for public is one of the main goals of many countries. For the last ten years, there have been number of works presented related to abnormal crowd behavior detection. The methods can be divided into two main groups that are holistic and non-holistic methods.

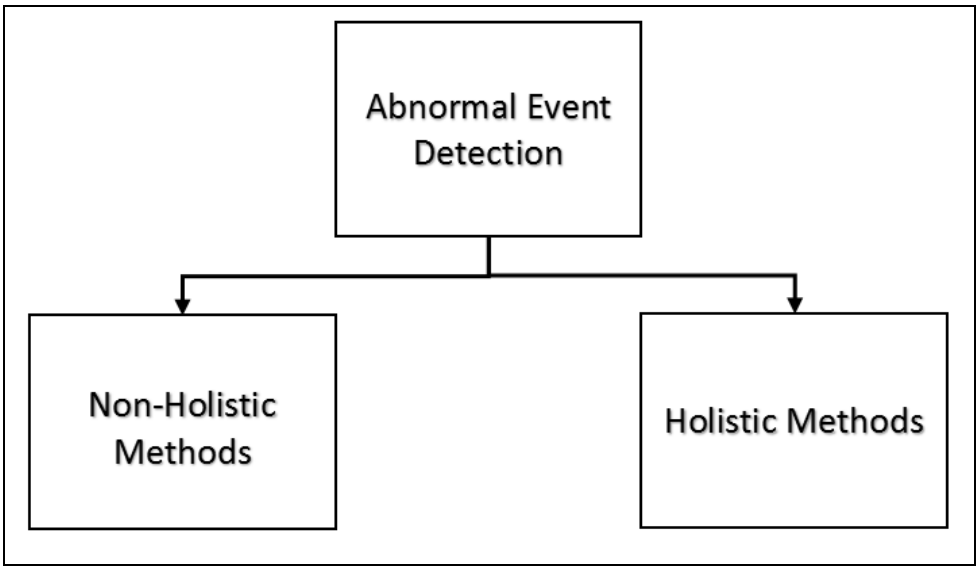

Figure 2.1: Main Branches of the approaches.

Non-holistic methods mainly focus on individuals on the scene in determining the abnormal human activities whereas holistic methods approach the scene as a whole entity instead of focusing each target. Since there are some difficulties in locating targets in crowded areas due to overlapping, Most of the contributions are made using holistic methods. Thus, we will be mainly discussing holistic methods that are done in estimating the abnormal crowd activities.

2.1. Non-Holistic Methods

Some researchers detect abnormal human activities based on individual behaviors of people in the scene. To do that Brostow et al. [2] proposes unsupervised Bayesian clustering algorithm to localize the individuals. Since occlusion in crowded areas leads background subtraction methods to fail when it comes to extract meaningful boundaries between individuals, Brostow tries to cluster points moving together assuming that they belong to same entity. In detection step, both Rusten-

Drummond [3] and Lukas-Kanade features are tracked by hierarchical optical flow algorithms. An illustration is given in Figure 2.2 below.

Figure 2.2: Feature Tracking of Brostow's method.

The Bayesian framework is utilized such that, assuming several points move together on each person in the scene. Brostow proposed a novel method to cluster most probable points given the distance matrix $Z(X_{i:N})$ which means to pick most likely clustering arrangement among M combinations.

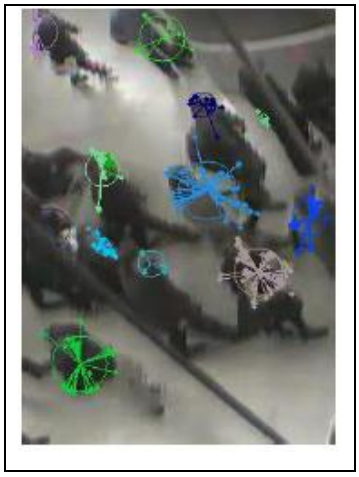

Figure 2.3: Brostow's Clustered Features.

Basharat et al. [4] tracks each object appears on foreground image obtained by background subtraction. A video yields m tracks $\{T_1, \dots T_m\}$ where each track consist of multiple observations of the same object such as $T_i = \{O_1, \dots O_n\}$ where $O_j = (t, x, y, w, h)$ denotes the observation j consisting of time stamp t, width w, height h, and the position of x, y information.

Motion patterns are modeled using the five dimensional random variable Γ_l for each pixel location where $\gamma = (x', y', \delta t, w_l, h_l)$ denotes one of the particular outcome of Γ_l. For each pixel location, multivariate gaussian mixture model is created which models the probability of that location being the source of transition. Probability of the observation γ belonging to GMM is given by

$$P(\Gamma_l = \gamma|\theta_l) = \sum_{i=1}^{n} \alpha_l^i p(\gamma|\theta_l^i) \quad (2.1)$$

Where n is the number of detected components in the mixture, θ_l^i is the parameters of the $i'th$ component $p(\gamma|\theta_l^i)$ has the gaussian pdf which can be expressed as

$$p(\gamma|\theta_l^i) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_l^i|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\gamma - \mu_l^i)^T \Sigma_l^{i-1}(\gamma - \mu_l^i)\right), \quad (2.2)$$

Where d is the dimension of the model with the parameters $\theta_l^i = \{\mu_l^i, \Sigma_l^i\}$.

2.2. Holistic Methods

Deciding if there is an abnormal situation in the crowd can also be done by analyzing the general behavior of the crowd rather than focusing on individual targets due to the fact that, people tend to exhibit 'herding behavior' in a dangerous case, which leads people to act together [5]. Various holistic approaches have been proposed for the sake of detecting abnormal behavior. As mentioned before, Non-holistic methods are inclined to fail where number of the individual is high in the scene because of overlapping [6]. Thus, considering the latest works in the literature, researchers have developed holistic methods to determine the crowd abnormality. Since the motion is one of the key parameter in determining the abnormal situation, Most of the researchers those propose holistic approach use an optical flow algorithm somewhere in their method. Because of that it is crucial to obtain accurate optical flow between consecutive frames to determine abnormalities. One of the famous approach is based on calculating social forces between particles moving by the optical flow as the time passes, first developed by Mehran et al. [5]. Later than some researchers developed enhanced social force models [7], [8]. Methods based on social force models are detailed in following part of the thesis.

2.2.1. Social Force Approach

This model is described to model human motion regarding some personal motivations and environmental limitations. Let say that person i changes his current velocity v_i such as

$$m_i \frac{dv_i}{dt} = F_a = F_p + F_{int} \tag{2.3}$$

where F_a, F_p F_{int} are actual force, personal desire force and interaction force. People in crowded sites often have desired locations to reach with the desired velocity v_i^p. But, due to congestion human motion is limited that cause a difference between actual velocity and desired velocity. For that reason people are inclined to reach their desired velocity based on personal desired force

$$F_p = \frac{1}{\tau}\left(v_i^p - v_i\right), \tag{2.4}$$

In Equation (2.4) τ is called "relaxation parameter".

F_{int} is composed of two forces F_{ped} which occurs due to the tendency of people to keep some distance from other people and the environment, F_w is the environmental force to prevent from hitting the walls or any obstacles. Since in a panic situation people tend to move together that is called "herding behaviors". In this kind of situation Mehran modified the Equation (2.4) by replacing v_i^p with

$$v_i^q = (1 - p_i)v_i^p + p_i\langle v_i^c\rangle, \tag{2.5}$$

In Equation (2.5), p_i,$\langle v_i^c\rangle$ denote the panic weight and average speed of surrounding pedestrians consecutively. This equation implies that as the panic weight increases velocity of a person approaches the average velocity of the people around him/her due to herding behavior. So, general formulation of social force model that Mehran proposed can be summarized as

$$m_i \frac{dv_i}{dt} = F_a = \frac{1}{\tau}\left(v_i^q - v_i\right) + F_{int} \tag{2.6}$$

Mehran treats particles as the member of the crowd and moves them using optical flow calculated between consecutive frames. It is stated that since the method

is not interested in object itself it is effective in low density sites as well as densely populated sites.

Calculating average optical flow in both time and space over a grid area is necessary to move particles on the image plane. To do averaging, Mehran uses Gaussian kernel filter in spatial domain. To initiate the process, particles are located over the image plane homogenously. For the sake of argument actual velocity is denoted by optical flow as

$$v_i = O_{ave}(x_i, y_i) \tag{2.7}$$

where $O_{ave}(x_i, y_i)$ is the averaged optical flow in both time and space domain for the particle i on the (x_i, y_i) coordinates. Then desired velocity can be given as

$$v_i^q = (1 - p_i)O(x_i, y_i) + p_i O_{ave}(x_i, y_i) \tag{2.8}$$

Linear interpolation method is used to calculate optical flow vector fields. To calculate the interaction forces of each particle Mehran proposed the Equation (2.9) given below which assumes that mass of each particle $m_i = 1$;

$$F_{int} = \frac{1}{\tau}\left(v_i^q - v_i\right) - \frac{dv_i}{dt} \tag{2.9}$$

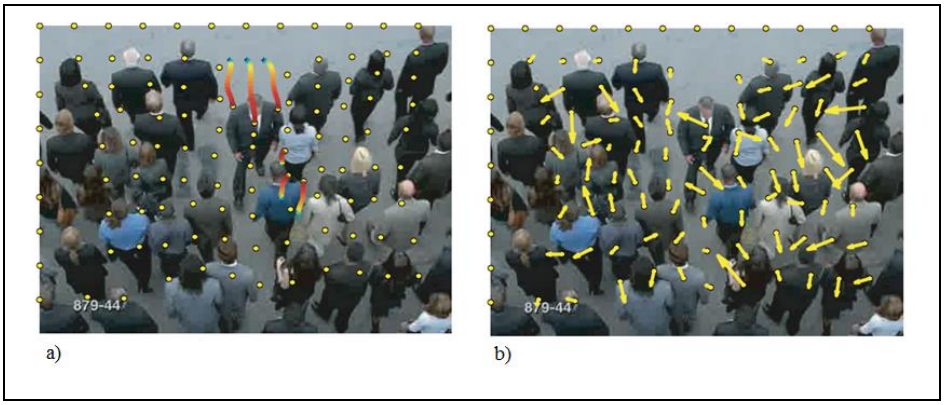

Figure 2.4: Particles interaction force demonstration. a) Particles, b) Force Vectors.

Determining whether video frame is abnormal or not is done by analyzing the pattern and the duration of the interaction forces rather than it's instantaneous value. In classification part, Mehran utilize bag of words approach where he used LDA (Latent Drichlet Allocation).

Zhang et al. [7], introduced a novel "Social Attributes-Aware Force" (SAFM) model in his work where he enhanced Mehran' social force model . In his method he defines interaction force as

$$F_{int}^{new} \propto W_{ij}^{S} \times \left(W_{ij}^{D} + W_{ij}^{C}\right) \times F_{int} \tag{2.10}$$

The term W_{ij}^{S} denotes scene scale estimation which is requied to quarantee that interaction force is convenient with the scene geometry. Zhang et al. [9] divide the image into the cells Γ_{ij} on which maximum and minimum scale $S_{\Gamma_{max}}$, $S_{\Gamma_{min}}$ and corresponding vertical coordinates $i_{\Gamma_{max}}, i_{\Gamma_{min}}$ are calculated. The Equation (2.11) is used to obtain scene scale weight

$$W_{ij}^{S} = (H - i) \times \left(\frac{S_{\Gamma_{max}} - S_{\Gamma_{min}}}{i_{\Gamma_{max}} - i_{min}} \times \frac{i_{\Gamma_{max}}}{i_{\Gamma_{min}}} - 1\right) / H + 1 \tag{2.11}$$

In Equation (2.10) W_{ij}^{D}, W_{ij}^{C} denote disorder attribute and congestion attribute respectively. In order to calculate these attributes low level motion features are required. $E_{disorder}$ is the force measure that represents disorder in the crowd. Similarly, $E_{congestion}$ is the model to define congestion behavior of the people. Both $E_{disorder}$ and $E_{congestion}$ formulations are given below:

$$W_{ij}^{D} = A_{ij} \exp\left(std(\varphi_{ij}) - std(\varphi_{T})\right) \tag{2.12}$$

$$W_{ij}^{C} = K_{ij} B_{ij} \left(\theta_{ij} - \theta_{T}\right) \tag{2.13}$$

where;

$$A_{ij} = sgn\left(std(\varphi_{ij}) - std(\varphi_{T})\right), \varphi_{ij} = Hist_{ij}\{O_n\}, n \in 1..8 \tag{2.14}$$

$$B_{ij} = sgn\left(\theta_{ij} - \theta_{T}\right), K_{ij} = std\left(Hist\left(V_{ij}\right)\right) \tag{2.15}$$

$std(.)$ denotes the standard deviation which is a good way to express changes of motion orientation that is denoted as orientation histogram with 8 bins φ_{ij}, $std(\varphi_T)$ is the threshold value, A_{ij} and B_{ij} are the signum functions. K_{ij} is the friction coefficient evaluated by standard deviation of the histogram of the optical flow.

Mehran's social force model has also been elevated by Zhao [8], who takes impact of the velocity field on the interaction forces into account such that along with the geometric position, probability of collision became another variable in calculating the force model which produces better results.

$$f_{i,j}^{int} = A_i e^{\frac{-d_{i,j}}{B_i}} \left(\lambda + (1-\lambda)\frac{1+\cos(\varphi_{i,j})}{2}\right)\left(\omega + (1-\omega)\frac{\cos(\theta_{i,j}+1)}{2}\right)\boldsymbol{v_{i,j}} \quad (2.16)$$

Zhao's interaction force model is given above, regarding this equation if the difference between velocity vectors of pedestrian i and pedestrian j interaction force would be small. $\boldsymbol{v_{i,j}}$ denotes the difference between $\boldsymbol{v_i}$ and $\boldsymbol{v_j}$ similarly $\theta_{i,j} = \arccos(\boldsymbol{v_i}, -\boldsymbol{v_{i,j}})$. Just like Mehran et al. Equally spaced particles which are assumed to represent pedestrians are used in this work. Particle's velocity is found using the Lukas-Kanade optical flow method [10]. To calculate the total interaction force on a particle Equation (2.17) is evaluated that denotes that particle $\boldsymbol{i}$ is effected by the particles within WxW square around it.

$$F_i^{int}(x_i, y_i) = \sum_{j \in win(W \times W)} f_{i,j}^{int} \quad (2.17)$$

For locating instability in the crowd, Zhao et al. [8] calculates the average optical and interaction forces within non-overlapping blocks and consider them as a feature vector $(\bar{V}, \bar{F})$. From this point K-means clustering to find K centroids. As the magnitude of the feature vector increases the crowd are more inclined to be instable meaning that pedestrians move faster.

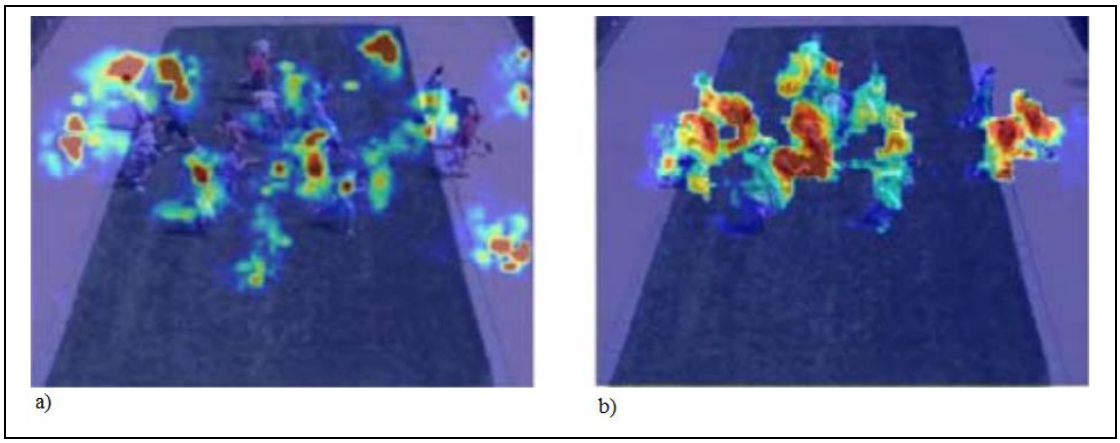

Figure 2.5: Comparison of interaction flows a) Mehran's method b) Zhao's method.

Abnormal behavior detection based on social force between equally spaced particles was first proposed by Mehran et al. [5]. Not requiring the detection of individuals and approaching the scene as a whole entity makes this method more robust for crowded scenes. Both Zhang [7] and Zhao [8] aimed to improve social force model, Zhang et al. Introduced social attribute-aware force model which regards social characteristics of the crowd whereas Zhao et al. proposed a novel velocity field approach. In Figure 2.6 framework of the algorithms based on social force are given.

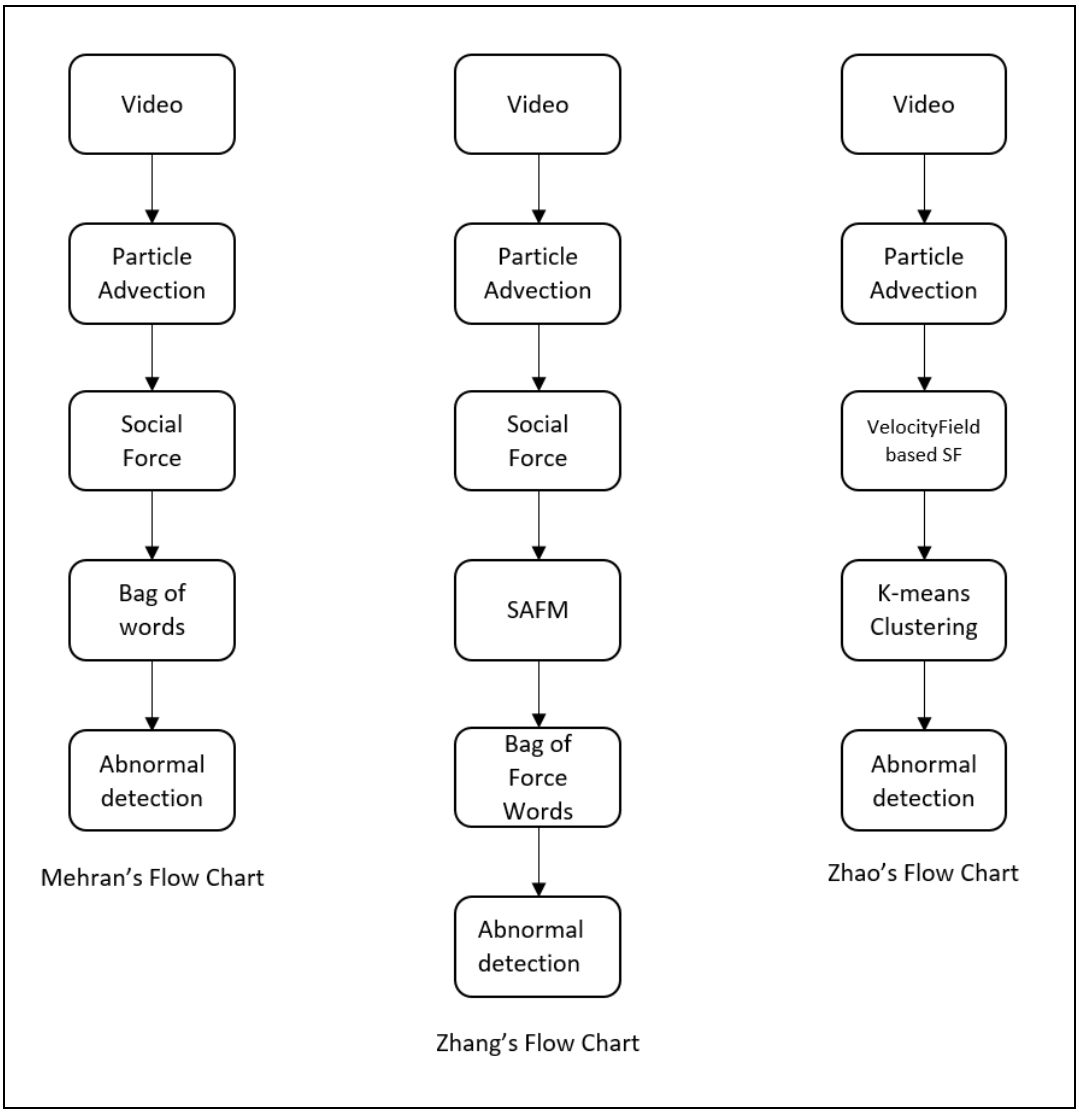

Figure 2.6: Algorithm steps of social force based methods.

2.2.2. Spatial-Temporal Feature Based Methods

Since abnormal events occur a period of time rather than a simple instantaneous changes between consecutive frames, both space and time related features can be used to determine the abnormal actions in the crowd. Du [9] estimates the likelihood of dynamic texture-motion representation called Structural Multi-Scale Motion Interrelated Patterns (SMMIP), which combines motion features and their structural spatial temporal information, to detect abnormal crowd behavior. Du utilizes Gaussian Mixture Model (GMM) to learn normal motion patterns and computes the likelihood estimation to decide if a patch is classified as abnormal.

In a $s_n \times s_n$ patch SSD (Sum of Squared Distances) are calculated for each pixel on the triplet frame. Suitability of motions are computed using the Equations (2.18), (2.19) given below.

$$D_1 = \sum_{m=1}^{s_n} \sum_{n=1}^{s_n} \left[I\left(m, n, t_p(1)\right) - I\left(m, n, t_p(2)\right) \right]^2 \qquad (2.18)$$

$$D_2 = \sum_{m=1}^{s_n} \sum_{n=1}^{s_n} \left[I\left(m, n, t_p(2)\right) - I\left(m, n, t_p(3)\right) \right]^2 \qquad (2.19)$$

D_1 and D_2 denotes the SSD scores. Important to point out that s_n and t_p denote different resolution of motion patterns.

$$S_{i,j}(\alpha) = \begin{cases} +1 & if\ D_1 - D_2 > \Theta \\ 0 & if\ |D_1 - D_2| \leq \Theta \\ +1 & if\ D_1 - D_2 < -\Theta \end{cases} \qquad (2.20)$$

After encoding each pixel by 8-trinary digit string per channel, positive and negative parts of the strings are separated. Collecting small patches negative and positive part of the code, 512-dimensional frequency histogram is created. Du [9] employ GMM to model normal motion patterns whose parameters are estimated by Expectation maximization. Furthermore, efficiency is enhanced by k-means method. To detect abnormalities the likelihood values L_p of patches are evaluated and compared to threshold T_p. If the L_p value is bigger than threshold T_p, the patch is classified as normal.

Local spatial-temporal motion patterns are also used by Kratz [11] to detect abnormalities. The distribution of spatial-temporal gradients are found for each pixel i in cuboid I. To find the spatio-temporal gradient ∇I_i following formula is used.

$$\nabla I_i = \left[I_{i,x}\, I_{i,y}\, I_{i,t} \right]^T = \left[\frac{\partial I}{\partial x}\ \frac{\partial I}{\partial y}\ \frac{\partial I}{\partial t} \right]^T \qquad (2.21)$$

where $x,\ y,\ t$ denotes the video's horizontal, vertical and temporal dimensions. For each pixel in a cuboid, a 3D Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is fitted to model the distribution of the gradients where

$$\boldsymbol{\mu} = \frac{1}{N}\sum_i^N \nabla I_i\,,\ \boldsymbol{\Sigma} = \frac{1}{N}\sum_i^N (\nabla I_i - \boldsymbol{\mu})(\nabla I_i - \boldsymbol{\mu})^T \tag{2.22}$$

The motion structure of the scene is captured by identifying the prototypical representations and extracting motion variations among the cuboids. Kullback – Leibler divergence method [11] is used to discriminate local motion patterns. For each motion pattern is represented by three dimensional Gaussian pdf. Kratz [11] decides if a new spatial-temporal cuboid O_t^n is a new prototype by measuring the KL distance between the spatial-temporal cuboid and known prototypes P_s . If the distances are greater than a threshold for all prototypes, cuboid is considered as a new prototype. Otherwise P_s is updated by the new observation O_t^n such that

$$P_s = \frac{1}{N_s+1} O_t^n + \left(1 - \frac{1}{N_s+1}\right) P_s \tag{2.23}$$

Kaltsa [12] uses a similar particle advection approach to Mehran uses. After placing symmetrical particles over the image plane,p denotes the position vector of each particle. To calculate the velocity of each particle she solves the Ordinary Differential Equation (ODE) (2.24) to interpolate flow field.

$$\dot{p} = \bar{v}(t, \bar{p}) \leftrightarrow \frac{d\bar{p}}{dt} = \bar{v}(t, \bar{p}) \tag{2.24}$$

$$\dot{p} = \bar{v}(t, \bar{p}), t \geq 0, \bar{p}(0) = \bar{p}_0 \tag{2.25}$$

Using the first order taylor series expansion Equation (2.26) can be expressed approximately as

$$\dot{p}(t_0 + h) \simeq \bar{p}(t_0) + h\frac{d\bar{p}}{dt}\Big|_{t=t_0} \tag{2.26}$$

Since each frame difference is equal to 1 in time domain, $h = 1$ is selected then Equation (2.26) can be expressed as

$$\hat{p}(t_0 + h) = \bar{p}(t_0) + \frac{d\bar{p}}{dt}\Big|_{t=t_0} = \left[\bar{p}(t_0) + \ \bar{v}(t, \bar{p})|_{t=t_0}\right] \tag{2.27}$$

Using the Equation (2.27) each particle on the image plane is moved frame by frame. Kaltsa extracts some features from each particle to cluster particles. Each particle at point $\bar{p}$, Kaltsa expresses its feature vector as

$$\bar{f}(\bar{p}) = [\bar{s}, \bar{p}, t_L, |\bar{v}|, \angle\bar{v}, n(\bar{p})] \tag{2.28}$$

where $\bar{s}$ is the parent source, $\bar{p}$ is the current position, t_L particle lifetime, $|\bar{v}|$, $\angle\bar{v}$ are magnitude and phase of the corresponding optical flow, $n(\bar{p})$ is the cluster id. Kaltsa uses DBSCAN algorithm [13] due to its noise robustness as well as not requiring initial number of clusters. Before detecting abnormality moving clusters direction are calculated using the formula:

$$\alpha = \arctan\left(\frac{1}{n}\sum_{j=1}^{n} sin\, a_j, \frac{1}{n}\sum_{j=1}^{n} cos\, a_j,\right) \tag{2.29}$$

In some cases arithmetic mean of vector angles give incorrect results Equation (2.29) is necessary in order to handle incorrect mean angle calculation.

Kaltsa measures the mean velocity of the all clusters then determine a threshold value. T_k is the test quantity obtained from each video frame which is then compared to threshold μ.

$$T_k = \sum_{i=1}^{N_c} \sum_{\bar{p}\in Cluster\ i} v_i(k, \bar{p}) + c \cdot \sigma \tag{2.30}$$

$$\eta = \mu \cdot N_{all} + c \cdot \sigma = \sum_{k=1}^{N_f} \sum_{i=1}^{N_c} \sum_{\bar{p}\in Cluster\ i} v_i(k, \bar{p}) + c \cdot \sigma \tag{2.31}$$

where

$$\mu = \frac{1}{N_{all}} \sum_{k=1}^{N_f} \sum_{\bar{p}\in Cluster\ i} v_i(k, \bar{p}) \tag{2.32}$$

$$\sigma = \frac{1}{N_{all}} \sum_{k=1}^{N_f} \sum_{\bar{p}\in Cluster\ i} v_i(k, \bar{p})^2 - \mu^2 \tag{2.33}$$

As the event behavior changes, T_k increases for all clusters. Kaltsa determines the dominant direction of the clusters via voting. The angle α of the cluster directions are categorized into one of the four basic direction up, down, left or right.

There are different methods based on optical flow proposed by other researchers. Wang [14] extracts KLT (Kanade-Lucas-Tomasi) corners and tracks them using optical flow. Since optical flow has a high computational cost, instead of generating optical flow of whole frame Wang only seeks corner points optical flow in order to reduce the computation time. Image plane is divided into blocks in which distribution of motion vectors itself then are modeled as a Gaussian distribution. The parameters of the Gaussian distribution $N(\mu, \sigma^2)$ are μ mean and σ^2 variance. Any given block motion pattern is described as $P(U, O)$ where U is the mean vector composed of mean velocity and mean direction μ_v, μ_r, O is the variance vector composed of variance velocity and variance direction σ_v^2, σ_r^2.

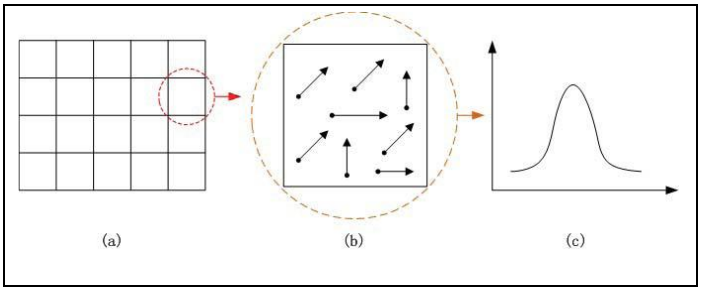

Figure 2.7: Motion descriptor. a) Grids, b) OF vectors, c) Histogram of Orientations.

Wang tries to cluster similar motion patterns $M(U, O)$. To do that, He uses a online clustering method which does not require initial number of clusters. First motion pattern is labeled as the first model then deviation between the first model and upcoming pattern is evaluated. If the deviation exceeds the threshold value new pattern is considered as a new model otherwise, it is assigned to a model which has the smallest deviation between them. The parameter update is done using the formula

$$M_k = \frac{1}{N_k+1} P_l + \left(1 - \frac{1}{N_k+1}\right) M_k \tag{2.34}$$

where M_k denotes pattern model, N_k is the number of the motion pattern, P_l is the upcoming pattern. Although overall accuracy of the proposed method of Wang [14] above %80 on some known datasets, there are some cons of the Wang method. Illumination and texture play important role when it comes to extract corners from image, Addition to that as the distance between people and camera increases less corner points represents moving objects that results less motion vectors. Not only the

corners cause some issues also in poor illuminated environment, traditional optical flow calculation produces inaccurate results. Wang also emphasizes that direction of walking effects optical flow since vertical movements towards to camera generates less motion vector whereas horizontal movements produces stronger motion vectors. Additionally, he emphasized that as the camera gets closer to the area, more accurate motion descriptor is captured which leads higher performance. Wang's test results given below are mostly based on UMN dataset.

2.3. Optical Flow

Motion is the key factor when it comes to detect abnormal behavior in the public areas. Thus it is important to calculate the motion as accurate as possible. Berthold K.P. Horn and Brian G. Schunk proposed a novel method called "Optical Flow" to calculate the motion vectors using two consecutive image frames in 1981 [15]. After their work was published, numerous optical flow algorithms [10] have been developed by many computer vision researchers. To do the algorithm, Horn and Schunk employed some constraints one of which is called "Brightness Constancy" a well known constraint in optical flow algorithms, other one is called "Smoothness Constraint". Details about these constraints will be given later on this chapter.

2.3.1. Traditional Models

In order to calculate optical flow Horn-Schunk assume that brightness intensity does not change at point $(x,\ y)$. If we denote image plane as $E(x, y, t)$ it can be seen that

$$\frac{dE}{dt} = 0 \tag{2.35}$$

Using the chain rule equation (2.35) can be expressed as

$$\frac{\partial E}{\partial x}\frac{dx}{dt} + \frac{\partial E}{\partial y}\frac{dy}{dt} + \frac{\partial E}{\partial t} = 0 \quad where \quad \frac{dx}{dt} = u,\ \frac{dy}{dt} = v \tag{2.36}$$

Substituting u and v Equation (2.36) becomes

$$E_x u + E_y v + E_t = 0 \tag{2.37}$$

where E_x, E_y and E_t are partial derivatives of the image with respect to x, y and t respectively. Equation (2.37) constitutes a line equation having infinite number of solution. Thus it is impossible to solve the Equation (2.37) for u and v without adding another constraint.

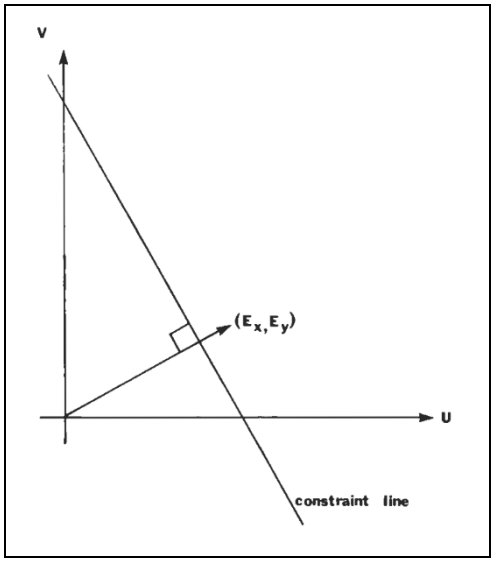

Figure 2.8: Optical flow constraint line.

If the points of the objects move independently, it would be hard to recover the velocities. As the opaque object moves, points closer to each other exhibit similar velocities, which means velocity field in the image should change smoothly. To express this additional constraint Horn-Schunk minimizes the square of the magnitude of the optical flow gradients.

$$E(u,v) = \iint \lambda(\, {u_x}^2 + {u_y}^2 + {v_x}^2 + \, {v_y}^2)^2 dxdy \qquad (2.38)$$

Horn-Schunk developed an iterative method to find u and v.

Bruce D. Lucas and Takeo Kanade [10] developed another widely used optical flow estimation method based on the idea that pixels in a small patch have the same flow, which enables brightness constancy equation to be solved by least squares criterion. Thus, local image flow vector V_x and V_y must satisfy

$$I_x(q_1)V_x + I_y(q_1)V_y = -I_t(q_1) \tag{2.39}$$

$$I_x(q_2)V_x + I_y(q_2)V_y = -I_t(q_2) \tag{2.40}$$

$$\vdots$$

$$I_x(q_n)V_x + I_y(q_n)V_y = -I_t(q_n) \tag{2.41}$$

where $q_1, q_2 \dots q_n$ are the pixels in the small patch and $I_x(q_i),\ I_y(q_i),\ I_t(q_i)$ are the partial derivatives of the image with respect to $x,\ y$ and t calculated at the pixel q_i. These equations can be respresented in matrix form $Av = b$ where;

$$A = \begin{bmatrix} I_x(q_1) & I_y(q_1) \\ I_x(q_2) & I_x(q_2) \\ \vdots & \vdots \\ I_x(q_n) & I_y(q_n) \end{bmatrix}, v = \begin{bmatrix} V_x \\ V_y \end{bmatrix} \text{ and } b = \begin{bmatrix} -I_t(q_1) \\ -I_t(q_2) \\ \vdots \\ -I_t(q_n) \end{bmatrix} \tag{2.42}$$

Since this system has more equations than the number of unknowns, it is over-determined. The least square principle is used to achieve a solution. To solve the 2x2 system following calculations are done.

$$A^T A v = A^T b \tag{2.43}$$

$$v = (A^T A)^{-1} A^T b \tag{2.44}$$

where A^T is the transpose of matrix

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum_i I_x(q_i)^2 & \sum_i I_x(q_i)\, I_y(q_i) \\ \sum_i I_y(q_i)\, I_x(q_i) & \sum_i I_y(q_i)^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i I_x(q_i)\, I_t(q_i) \\ -\sum_i I_y(q_i)\, I_t(q_i) \end{bmatrix} \tag{2.45}$$

In order to obtain this solution, there are some conditions that should be met. First, $A^T A$ should be invertable also its value should not be too small because of noise. $A^T A$ must also be well-conditioned. However, there are some potential errors that can be encountered in calculating the traditional optical flow (Horn-Schunk or Lucas-Kanade). Since brighness constancy and small motion assumptions can be

violated easily in real world situations, the optical flow could often produce unreliable results.

After Horn-Schunk, many computer vision researchers proposed novel methods that could produce more realistic flow vectors [16], [17], [18]. The newest approach that outperforms all the methods from the literature so far is the optical flow estimation based on a theory for warping represented by Thomas Brox [17] in 2004. He introduced a novel variational model containing multiple constraints which will be discussed later on this part.

2.3.2. Variational Model

In traditional optical flow, Equation (2.37) is obtained by using linearized taylor series expansion of the brightness constancy equation which can actually work only in the condition that, displacement vector is small. In many cases this small displacement assumption is violated. To overcome this issue Brox [17] does not linearize the brightness constancy equation as well as gradient constancy equation. They combine non-linear brightness and gradient constancy constraint as a data energy function to be minimized.

$$E_{Data}(u,v) = \int_{\Omega} \psi(|I(\vec{x}+\vec{w}) - I(\vec{x})|^2 + \gamma|\nabla I(\vec{x}+\vec{w}) - \nabla I(\vec{x})|^2)d\vec{x} \quad (2.46)$$

Ω ' denotes the image domain where $E_{Data}(u,v)$ energy function to be integrated over. Where $\vec{x} = (x,y,t)^T$ and $\vec{w} = (u,v,1)^T$. $I(\vec{x}+\vec{w}) - I(\vec{x})$ is the non-linearized gray value constancy whereas $\nabla I(\vec{x}+\vec{w}) - \nabla I(\vec{x})$ part is the gradient constancy assumptions. Addition to these assumptions, Brox also takes smoothness assumption into account such that;

$$E_{Smooth}(u,v) = \int_{\Omega} \psi(|\nabla_3 u|^2 + |\nabla_3 v|^2)d\vec{x} \quad (2.47)$$

where $\nabla_3 = (\partial x, \partial y, \partial t)^T$, the function $\psi(s^2) = \sqrt{s^2+\epsilon^2}$ due to the small constant convexity is maintained which offers advantages in minimization process without introducing another parameter rather than constant ϵ which is fixed to be 0.001. Total function to be minimised is the weighted sum of E_{Data} and E_{Smooth} terms such that

$$E(u,v) = E_{Data}(u,v) + \alpha E_{Smooth}(u,v) \tag{2.48}$$

α is called the regularization parameter which is greater than 0. From this point the purpose is to find u, v that minimises energy function (u, v) . The calculus of variations states that minimising functions must fulfil the Euler-Lagrange Equations such that:

$$\begin{aligned} &\psi'\left(I_z^2 + \gamma\left(I_{xz}^2 + I_{yz}^2\right)\right).\left(I_x I_z + \gamma\left(I_{xx}I_{xz} + I_{xy}I_{yz}\right)\right) \\ &-\alpha\, div(\psi'(|\nabla_3 u|^2 + |\nabla_3 v|^2)\nabla_3 u) = 0, \end{aligned} \tag{2.49}$$

$$\begin{aligned} &\psi'\left(I_z^2 + \gamma\left(I_{xz}^2 + I_{yz}^2\right)\right).\left(I_y I_z + \gamma\left(I_{yy}I_{yz} + I_{xy}I_{xz}\right)\right) \\ &-\alpha\, div\left(\psi'(|\nabla_3 u|^2 + |\nabla_3 v|^2)\nabla_3 u\right) = 0, \end{aligned} \tag{2.50}$$

In order to solve this problem numerically Brox uses fixed point iterations on w along with downsampling method to obtain better approximate global optimum energy. Unlike using traditional 0.5 down sampling factor on each level arbitrary sampling factor is used in his method which yields smoother transition between scales. Mathematical details that can be examined in the original paper will not be covered in this thesis. Evaluating the performance, famous image sequence "*Yosemite*" with and without cloud was used by Brox. We also share the angular error table given in the Brox's paper [17].

Table 2.1: Yosemite sequences performance results of Brox compared with other methods. AAE = average angular error. STD =standard deviation

Yosemite with cloud			**Yosemite without clouds**		
Technique	**AAE**	**STD**	**Technique**	**AAE**	**STD**
Nagel[19]	10.22°	16.51°	Ju et al[20]	2.16°	2.00°
Horn-Schunk[15]	9.78°	16.19°	Bab-Hadiashar-Suter[21]	2.05°	2.92°
Uras et al.[19]	8.94°	15.61°	Lai-Vemuri[22]	1.99°	1.41°
Alvarez et al[23]	5.53°	7.40°	Brox(2D)[17]	1.59°	1.39°
Weichkert et al[24]	5.18°	8.68°	Memin-Perez[25]	1.58°	1.21°
Memin – Perez[23]	4.69°	6.89°	Weickert et al[24]	1.46°	1.50°
Brox(2D)[17]	2.46°	7.31°	Farneback[26]	1.14°	2.14°
Brox(3D)[17]	1.94°	6.02°	Brox(3D)[17]	0.98°	1.17°

Performance of Brox's optical flow based on warping theory outperforms other methods in the literature. Since motion is one of the key factor in detecting the abnormal crowd behavior, we decided to utilize this algorithm due to it's reliability and robustness in handling large displacements.

3. PROPOSED METHOD

In Section 2, we have mentioned abnormal crowd behavior detection methods based on various techniques. As many researchers those have studied this problem stressed that holistic approaches are more suitable for detecting abnormal actions especially in crowded scenes [5], [6], [9], [14]. There are two main reasons; the first one is that, in a crowded area, it is difficult to detect individuals. Secondly, due to overlapping, it is a challenging task to track individuals in order to extract their paths. Having analyzed the previous works in the literature, it is seen that both space and time information possess hints when it comes to understand the crowd behavior when unusual events occur. For instance, when people react to something, optical flow of the image plane changes due to the human motion. Besides, duration of the changes in spatial domain is also important feature to consider since abrupt changes does not necessarily mean that there is an abnormal situation.

We are inspired by the idea that how the motion information is effective in determining the abnormal crowd activities. We have seen that some researchers [14], [27], [5], [6], [12] utilize traditional dense and sparse optical flow algorithms such as Horn-Schunk [15] and Lucas-Kanade [10]. As we discussed in previous chapter various optical flow algorithms have been developed to represent the motion accurately. As far as performances are concerned, Brox's optical flow method [17] outperforms all other methods in the literature. Thus, it is decided to use his algorithm to calculate the motion information. Test results will be given in last section of the thesis.

After calculating the optical flow, one needs to detect the abnormal events from this motion information. To classify the events as normal or abnormal, we have followed a similar motion influence matrix procedure that Lee at al presented his paper [6] we found that lee's influence matrix method has powerful sides and capture the abnormal events yet requires some modifications. Later in Section 4 we give test results of our proposed method on various datasets.

3.1. Grid Based Approach

In our method, we divide frames into grids and each grid A_i where $i = 0..L$ is an object that has some properties which are used to calculate the influence map for each frame.

$$Frame = [A_1, A_2, .. A_L] \tag{3.1}$$

$$A_{Opt} = [V_1(u,v), V_2(u,v) \dots V_{MxN}(u,v)] \tag{3.2}$$

where A_{Opt} corresponds grid's optical flow property, V corresponds to optical flow vector that has two components.

The properties of each grid are given below:

- Position
- Average Magnitude
- Dominant Direction
- Influence
- Influence History

For each frame we update these properties based on the optical flow calculation matrixes. Unlike Lee's method [6], calculation of average magnitude and dominant direction are done using the Equations (3.3) and (3.4) given below.

$$A_{mag} = \rho_A = [\sum_{i \in A} u_i + \sum_{i \in A} v_i]^{1/2} \tag{3.3}$$

$$A_{dir} = \theta_A = arcta\, n\left(\frac{-\sum_{i \in A} u_i}{\sum_{i \in A} v_i}\right) \tag{3.4}$$

One of the difference between our method and Lee et al. [6] is that, he quantize the orientation of the optical flow magnitude of a grid into 8 bins while our method does not quantize the orientation angle. More details about the differences between our method and Lee's method will be given in the following Influence Map section.

3.2. Influence Map

Grid based approach allow us to localize abnormalities based on obtained properties using high performance optical flow method within these grids. As we think of an abnormal situation from optical flow perspective, direction and magnitude of each grids optical flow matrixes should be taken into account. Influence map is created regarding these grid's optical flow properties such as magnitude, direction and the duration. Basically, each grid object has influence property updated by the formulas (3.5), (3.6), (3.7) given below. As grid influence value increases, it is likely that an abnormal event has occurred in corresponding grid location.

$$w_{ij} = w^{d}_{A_iA_j}\, \varphi_{A_iA_j} \exp\left(\frac{-D_{A_iA_j}{}^{2}}{\rho_i}\right) \tag{3.5}$$

$$w^{d}_{A_iA_j} = \begin{cases} 1, & D_{A_iA_j} < R_{influence} \times \rho_i \\ 0, & other \end{cases} \tag{3.6}$$

$$\varphi_{A_iA_j} = \begin{cases} 1, & Abs(\angle\rho_i - \angle A_iA_j) < \pi/2 \\ 0, & other \end{cases} \tag{3.7}$$

Influence value of a grid is generated by other neighboring grid's optical flow matrices. Such that w_{ij} means influence value of grid A_j generated by grid A_i. $D_{A_iA_j}$ denotes the distance between grid A_i and grid A_j. There is an inverse non-linear relation between influence value and distance between grids. $R_{influence}$ parameter determines the base influence range of each grid. $w^{d}_{A_iA_j}$ denotes the influence range coefficient of A_i such that if $R_{influence} \times \rho_i$ value is less than the the distance between grids then grid A_i does not influence grid A_j. Unlike Lee et al. [6], who does not include ρ_i value in calculating $w^{d}_{A_iA_j}$, In our algorithm, if a grid's average optical flow is high then its influence range is also wide. $\varphi_{A_iA_j}$ denotes the angle coefficient between grid A_i and grid A_j if the absolute difference between dominant direction angle and the angle of line that connect the center of two grids is less than π then the grid A_i influence grid A_j. Purpose of using the direction of grids optical flow vectors

is to represent location of abnormality more accurate. The total influence energy $E(j)$ that a grid has is the sum of all the influence values that are generated by other grids.

$$E(j) = \sum_{i=1}^{n} w_{ij}\, \rho_i \qquad (3.3)$$

In Figure 3.1 influence maps of some sample frames are illustrated. It can be seen that as people move in a hurry, location of where action happens becomes hotter in influence map indicating that abnormal event is happening.

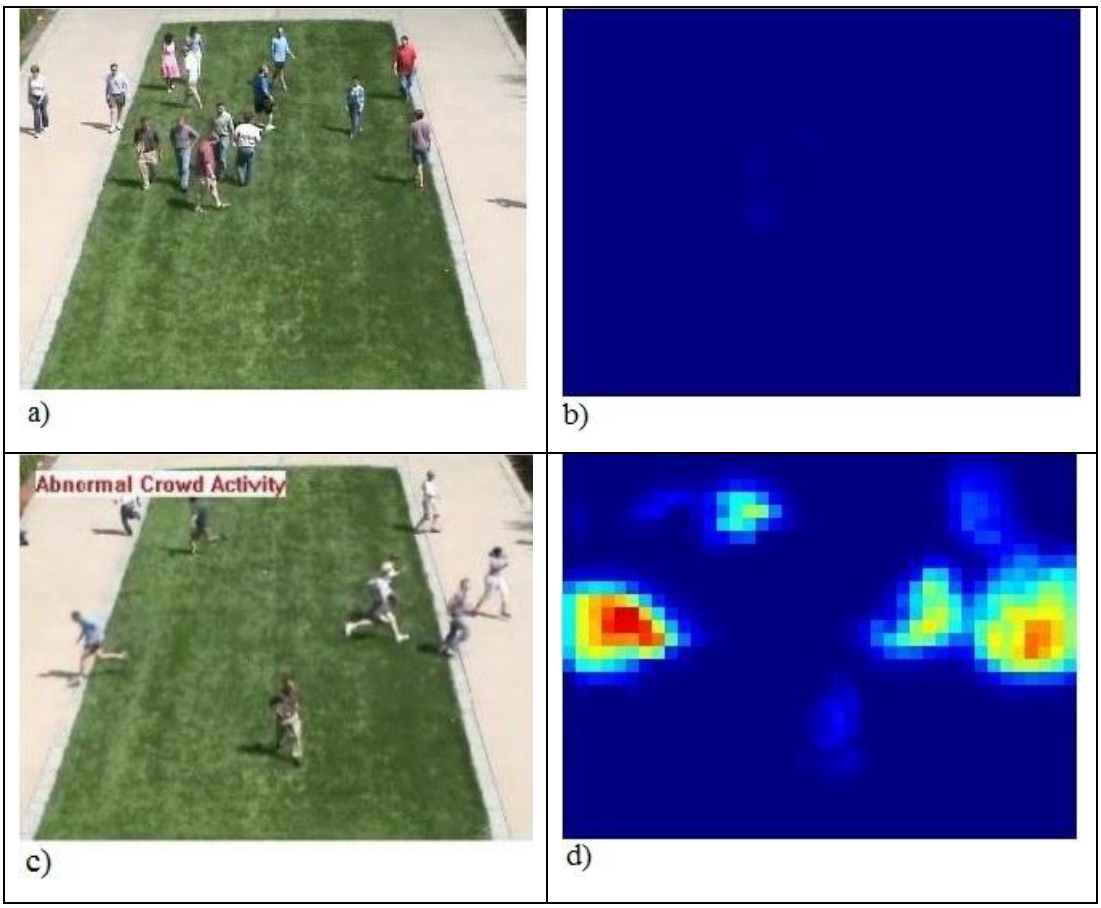

Figure 3.1: a) Normal Frame, b) Normal Influence map, c) Abnormal frame, d) Abnormal Influence map.

3.3. Framework of The Algorithm

Our proposed method basically is composed of six stages giving in Figure 3.2 Details of each stage is discussed in following portion of this part.

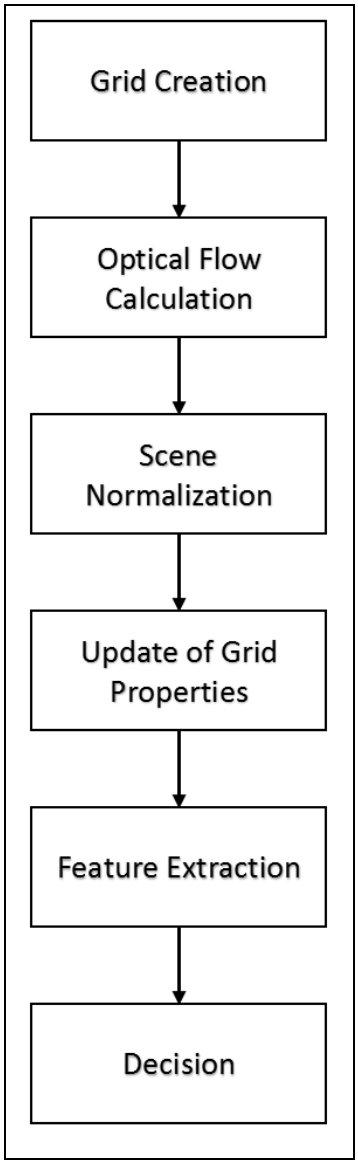

Figure 3.2: Framework of the proposed method.

3.3.1. Grid Creation

Rather than using image plane directly, we utilize grid objects to represent motion and detect abnormal situation. Initially only argument that grid object requires is the center of window location. We locate those grids considering the

equally spaced grids over an image plane so that distance between each neighboring grid is equal and whole grids constitute a rectangular area on the image plane.

3.3.2. High Performance Optical Flow Calculation

Since the motion is the fundamental input of our method, accurate optical flow information is necessary to create reliable influence map. We have performed Brox's high performance optical flow algorithm which outperforms all other optical flow algorithms on the literature [17]. To demonstrate the difference between traditional and Brox's optical flow algorithms, an example is given below

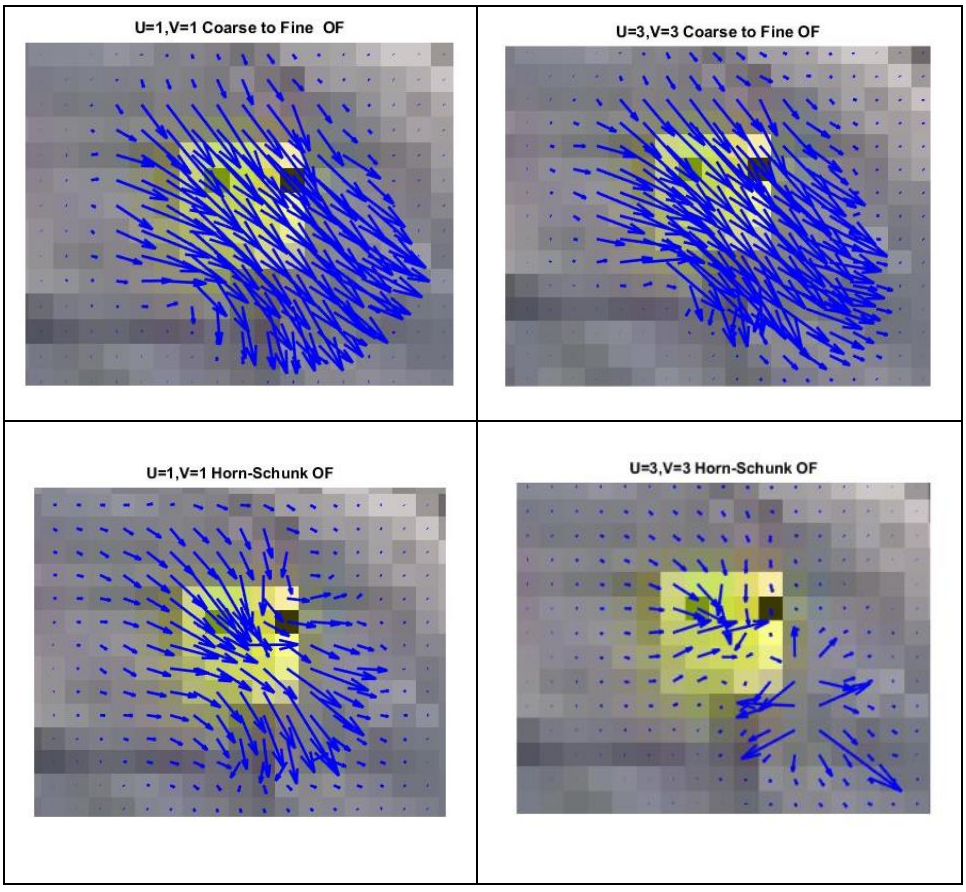

Figure 3.3: The Comparison between Horn-Schunk (bottom row) and Thomas Brox (top row) optical flow.

In Figure 3.3, we have demonstrated OF (optical flow) of Horn-Schunk and the Brox's method. The yellow square is moved 1 and 3 pixels in both x and y direction respectively and the optical flows of each case is calculated. Since traditional optical flow approach uses linearized gray value constancy equation (3.37), Horn-Schunk

optical flow calculation fails as the displacement increases. On the other hand, yielding reliable flow field, Brox's high performance optical flow based on the warping theory can handle both cases due to its robustness to large displacement.

3.3.3. Scene Normalization

Obtained motion information can vary due to the distance between camera and the object/person within the scene. Because of that, even if two people walking with the same speed, generated optical flow vectors will be different due to the different distances of each individual to the camera [14]. This situation could be an issue where obtained average optical flow vector of a person, who is closer to camera, is similar as the flow vectors obtained from a person running far away from the camera. To balance these vectors we create a normalization matrix and multiply it by the U and V matrices coming from optical flow calculation. This normalization can be considered as an approximate approach since we only regard scene middle furthest – nearest points to create normalization matrix. Also we utilize this approach only in GTU dataset since the scene properties of UMN dataset is unknown. In Figure 3.4, a sample camera scene is depicted. Although both of them have the same velocity, Because of their distance to camera the optical flow vector generated from the person closer to camera is greater (yellow arrow) than the vector (blue arrow) generated from the person located further away to camera.

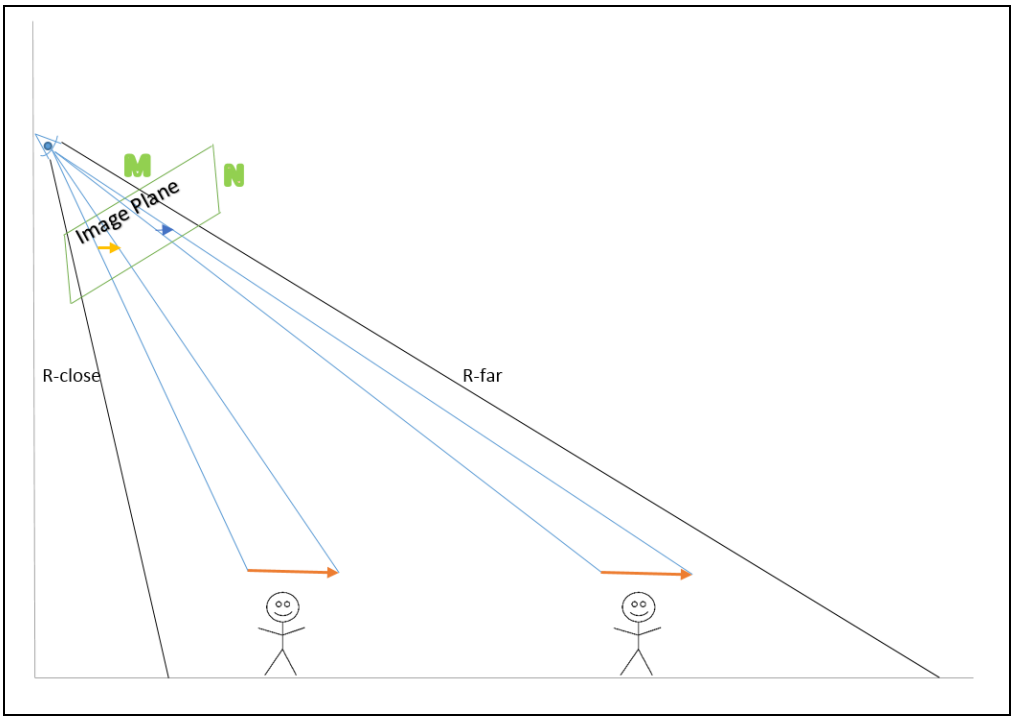

Figure 3.4: Demonstration of scene and image plane

$$UB = \frac{R\_far}{R\_close} \tag{3.9}$$

$$LB = 1 \tag{3.10}$$

$$NormalizingMatrix = \begin{bmatrix} UB & UB & UB & UB & UB \\ LB + (M-2)\frac{UB-LB}{M-1} & LB + (M-2)\frac{UB-LB}{M-1} & \dots & \dots & LB + (M-2)\frac{UB-LB}{M-1} \\ \vdots & \vdots & \dots & \dots & \vdots \\ LB + 2\frac{UB-LB}{M-1} & LB + 2\frac{UB-LB}{M-1} & \dots & \dots & LB + 2\frac{UB-LB}{M-1} \\ LB + \frac{UB-LB}{M-1} & LB + \frac{UB-LB}{M} & \dots & \dots & LB + \frac{UB-LB}{M-1} \\ LB & LB & LB & LB & LB \end{bmatrix}_{MxN} \tag{3.11}$$

3.3.4. Update of Grid Properties

As the time passes we update grid properties using the optical flow information. In order to calculate the influence values between grids, Euclidean distance information between corresponding grids is required. This matrix is later used to find the grids effected by for each particular grid. $DM(i,j)$ is the symmetrical distance matrix which indicates the distances between grid i and j. Each distance

$D_{A_iA_j}$ is calculated by using the Equation (3.12) and $DM(i,j)$ matrix is formed by these $D_{A_iA_j}$ values. To update the grid optical flow properties, we calculate the optical flow of consequent frames for every time instance and by using the Equations (3.3), (3.4) each grid optical flow properties are updated. These properties are necessary to calculate influence values for each grid.

$$D_{A_iA_j} = \sqrt{\left(A_{i_x} - A_{j_x}\right)^2 + \left(A_{i_y} - A_{j_y}\right)^2} \tag{3.12}$$

.

$$DM(i,j) = \begin{bmatrix} 0 & D_{12} & D_{13} & \cdots & D_{M1} \\ D_{21} & 0 & \cdots & \cdots & \\ D_{31} & \cdots & 0 & \cdots & \\ \vdots & \cdots & \cdots & \ddots & \vdots \\ D_{M1} & \cdots & \cdots & \cdots & 0 \end{bmatrix} \tag{3.13}$$

After update of grid optical flow properties, we update the influence values of each grid. To do that we first decide some parameters such as $R_{influence}$, $R_{feature}$ and $R_{temporal}$. $R_{influence}$ determines the range of the grid's influence area while $R_{feature}$ determines the radius of feature extracting area and $R_{temporal}$ is required to take the duration of an action into account. Values of these parameters are decided based on some performance evaluations on training dataset. Considering these experiments we arranged parameters as $R_{influence} = 10$, $R_{feature} = 30$, $R_{temporal} = 8$. In Figure 3.5, an example influence map formed by a single grid is demonstrated. The black dotted grid with the arrow is the influencing grid whereas the grids with the blue, yellow and red colors are the influenced grids from black one. Considering the direction of the grid's optical flow property, grids with red color are the most influenced grids while the blue ones are the least influenced grids.

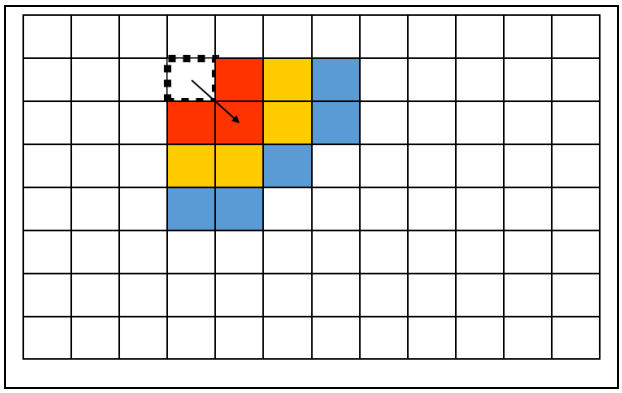

Figure 3.5: Influence map of single grid.

3.3.5. Feature Extraction

Since our ultimate goal is to decide if a video frame is normal or not, we try to decide a threshold for each scene. As a feature, we define 'Scene Energy Value (SEV)' considering the grid that has the maximum cumulated influence value. As an abnormal event has a duration, for each frame the grid that has maximum influence value is found and it's temporal influence average value is considered as scene energy value.

$$SEV(t) = \frac{\sum_{\tau=-\frac{R_{temporal}}{2}}^{\tau=+\frac{R_{temporal}}{2}} \left(Max(A_i)(t-\tau)\right)}{R_{temporal}+1} \tag{3.14}$$

In Figure 3.6, scene energy graph is demonstrated. Orange line denotes the threshold value obtained by training clips of particular scene. Black vertical line separates Normal and Abnormal frames. It could be seen that energy of frames in normal area is low due to the normal motion of individuals. After a time people start running in panic which leads extracted frame energy to be greater than the threshold value. Our method is powerful and produce high detection rate especially in crowd scenes since each grid energy is effected by its neighboring grids influence values. As more people move together the accumulated grid energy increases.

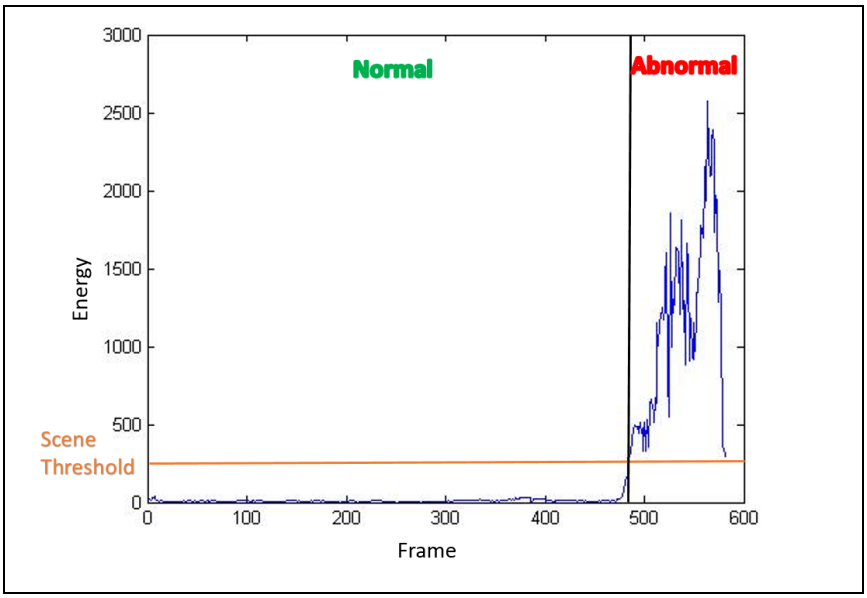

Figure 3.6: Scene Energy Graph.

3.3.6. Decision

Our performance criteria is based on one dimensional feature value for each frame. To get the feature value, as mentioned in Section 3.3.5, Grids Influence Values are extracted using both spatial and temporal information. In Section 4.1.1 and 4.2.1 important frame numbers are given for both UMN and GTU dataset. For each video scene, one threshold value is calculated using the training clips. To evaluate the threshold value, we iteratively increment T from 0 to 1000 and for each step. Calculating the correctly and incorrectly labeled frames, an accuracy ratio for each T value is found. Then the T value which gives the lowest error is selected as Threshold value.

$$Threshold = argmin\big(Error(T)\big) \tag{3.15}$$

Decision of whether or not a frame is normal or abnormal, the corresponding $SEV(t)$ of frame, is compared with scene threshold value. If the SEV is greater than the scene threshold, the frame is classified as abnormal otherwise it is classified as normal. We use “leave one out” approach when calculating the accuracy of the algorithm. Flowchart of the decision rule is given in the Figure 3.7.

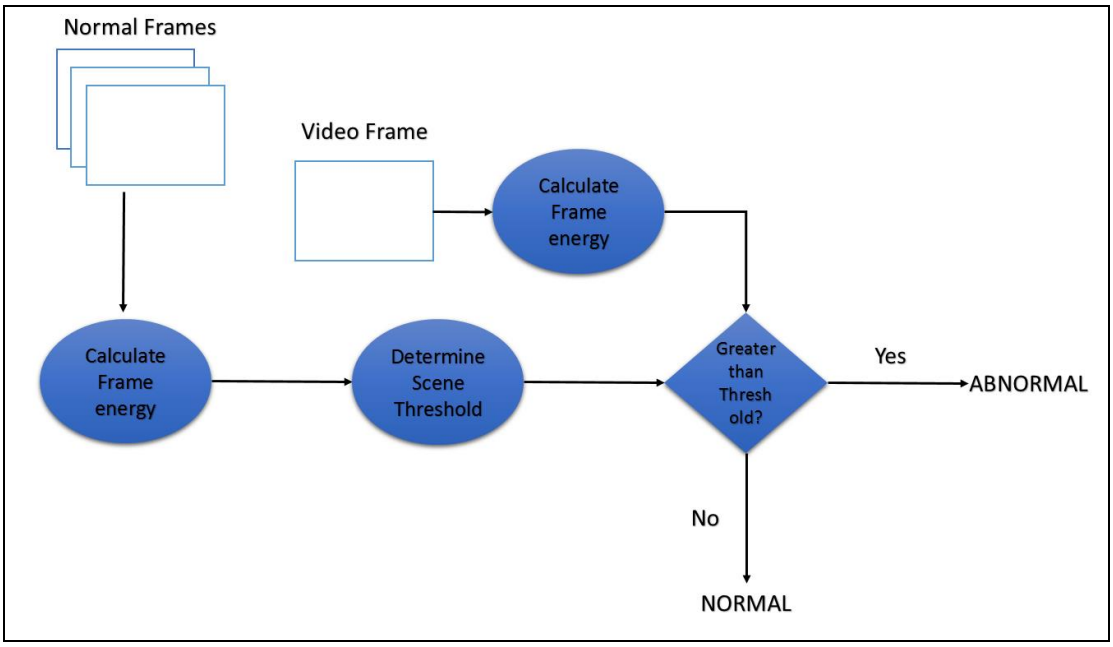

Figure 3.7: Flowchart of the decision mechanism.

3.4. Determination of Threshold

In our approach we represent each frame by a single feature value which is called 'Scene Energy Value-SEV' (details about obtaining SEV value are given in Section 3.3.5). How we determine the scene threshold value is critical. We use two different methods to calculate this threshold. The first method is Brute-Force Threshold (BFT) method which requires both abnormal and normal information during the training, which we call Normal Frame Thresholding (NFT). The other method is based on finding the maximum SEV value of normal training clips. To evaluate the performance of the proposed algorithm, we use leave-one-out cross validation method as the number of video clips is small. We use the clips belonging to the same scene when determining the threshold. Clips with different scene are not used together when applying the leave-one-out method. Pseudo codes of each thresholding techniques are given below.

- Brute-Force Thresholding(BFT)

  -Initiate variables

  -True_Classified = 0;

  -False_Classified = 0;

  -Accuracy=0;

-OptimumT=0;

-For T from 0 to a large value

-Calculate True and False Classified Frames

- Calculate Accuracy

- If Accuracy is greater than the Accuracy for previous T, Set T to optimumT

-End For

-Calculate Accuracy for optimum

- Normal Frame Thresholding (NFT)

-Take Training Normal frames;

-Assign T to average of $\sum_{i \in M}$ Max(SEV(n)), where n $\in$ $\forall$ Normal Training

-Frames, M denotes number of training clips;

-Return T;

4. TEST

4.1. Datasets

We have used two different datasets to evaluate our algorithm's performance. First dataset was created by University of Minnesota and the second one was created by Gebze Technical University. Each dataset contains a group of people where they walk at the beginning of each clip then, they start running due to an unexpected event.

4.1.1. UMN Dataset

This dataset is composed of nine clips with three different scenes. The resolution of the clips is 240x320 pixels. In Figure 4.1, there is an illustration of normal and abnormal sample frames. The left column shows the normal frames and the right column shows the abnormal frames. Each row gives representative frames from the three scenes. The first scene consists of two different video clips captured in a sunny day. The second scene contains five different clips with different escape scenarios. Unlike the first and the third scene, the second scene is from an indoor environment and illuminated poorly. The third scene has two clips in an outdoor environment with an adequate illumination. When comparing the scenes in terms of crowdedness, one will notice that all of the video clips possess similar crowdedness, from 10 to 20 people for each scene. At the beginning of each video, people exhibit regular behaviors such as walking or talking to each other. After a time, people start running suddenly to evacuate the area.

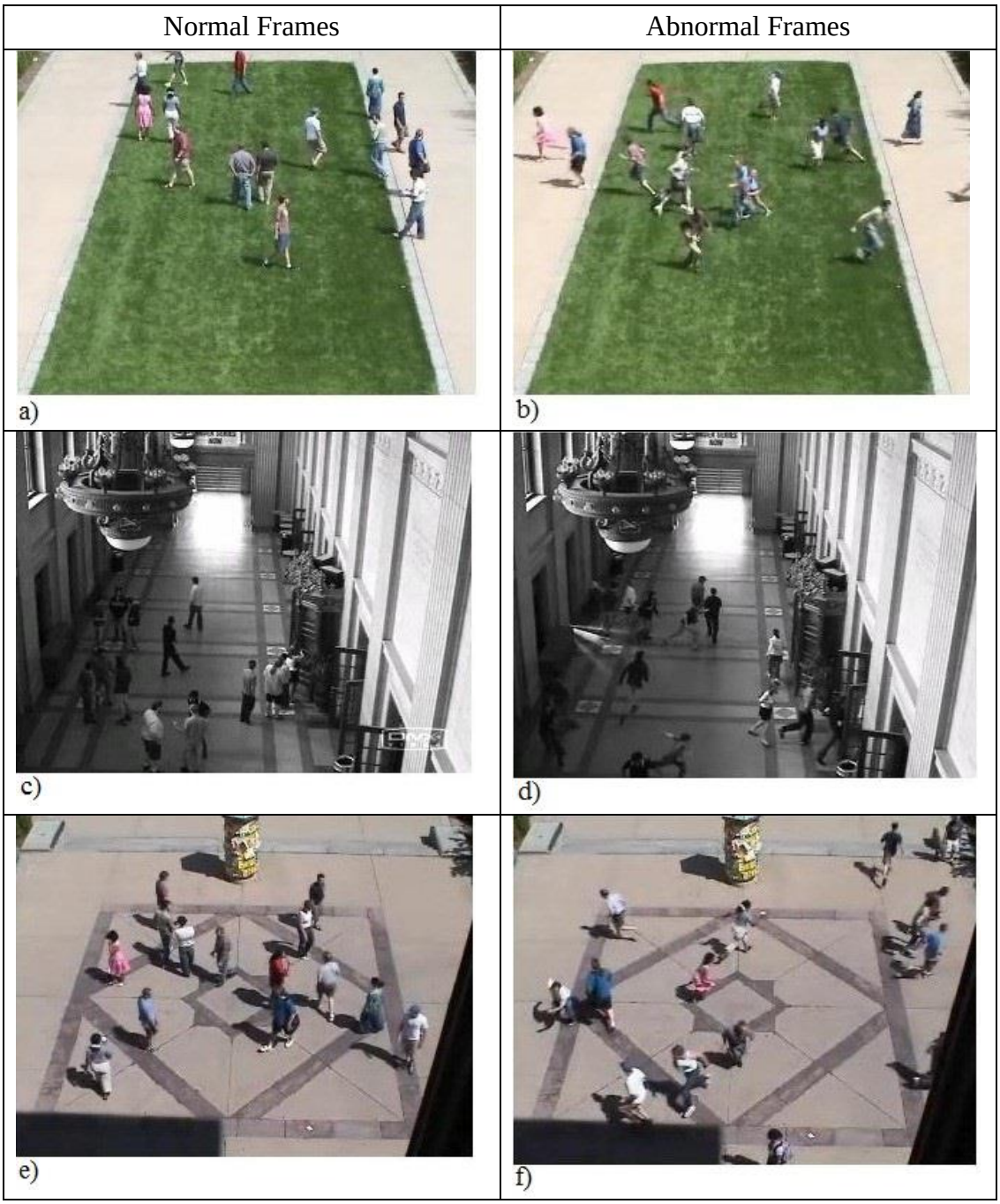

Figure 4.1: Sample frames of abnormal and normal frames. a) UMN1 normal, b) UMN1 abnormal, c) UMN2 normal, d) UMN2 abnormal, e) UMN3 normal, f) UMN3 abnormal.

- Ground truth

UMN dataset does not possess ground truth information. Since the term 'abnormal' can be subjective from person to person, it is not certain that which frame is the beginning of the abnormal event in the video frames. Therefore, we need to

make a judgment when determining the ground truth. We have compared the related works in the literature in terms of the ground truth. Since UMN dataset only involves GAE, we decided to choose the initial frame as the beginning the abnormal event when half or more of the people in the scene start running. This approach provides us a similar ground truth as compared to the ground truth given graphically in other studies. The UMN dataset consist of 7739 frames. In Table 4.1, 4.2 and 4.3, the ground truth information is given. In the first column, NF denotes Normal Frames while AF denotes Abnormal Frames.

Table 4.1: Scene-1 Ground truth.

UMN Dataset	**Sc:1 Clip No:1**	**Sc:1 Clip No:2**
NF start	1	893
NF stop, AF start	480	1308
AF stop	590	1366

Table 4.2: Scene-2 Ground truth.

UMN Dataset	**Sc:2 Clip No:1**	**Sc:2 Clip No:2**	**Sc:2 Clip No:3**	**Sc:2 Clip No:4**	**Sc:2 Clip No:5**
NF start	2038	2714	3592	4062	4952
NF stop, AF start	2580	3177	3927	4776	5392
AF stop	2665	3283	3988	4881	5505

Table 4.3: Scene-3 Ground truth.

UMN dataset	**Sc:3 Clip No:1**	**Sc:3 Clip No:2**
NF start	5625	6255
NF stop, AF start	6144	6835
AF stop	6226	6900

4.1.2. GTU Dataset

In order to validate our method's reliability we need additional test videos. For this purpose we have recorded some surveillance videos considering different abnormal situations. Unlike UMN dataset we also took local abnormal scenarios into account such that, only single or two people behave abnormal. Crowdedness of the scene is similar to UMN dataset where between 15 to 20 people are present on the scene. Dataset is composed of two different scenes with 320x240 pixel-sized video clips some of which do not include any abnormal events.

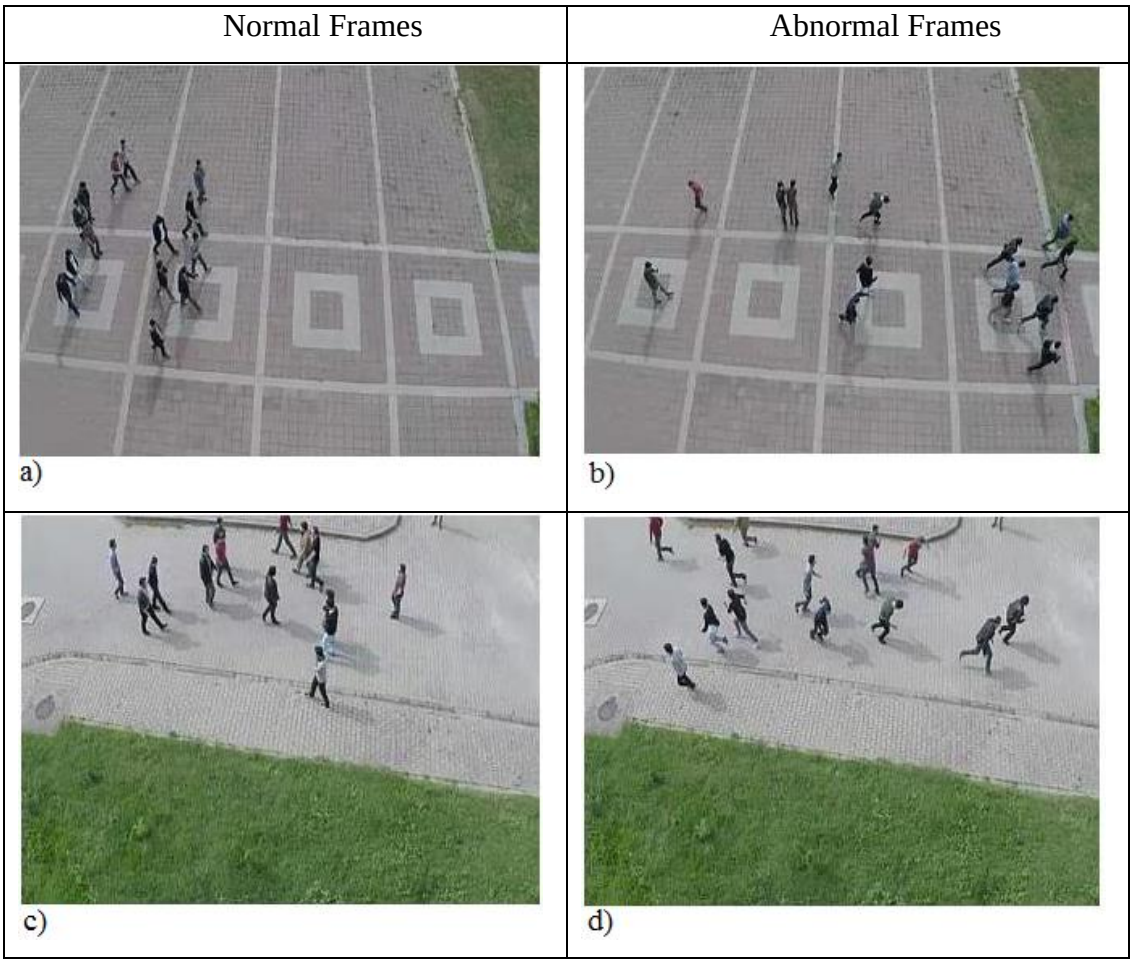

Figure 4.2: GTU Dataset sample abnormal and normal frames. a) GTU1 normal, b) GTU1 abnormal, c) GTU2 normal, d) GTU2 abnormal.

- **Ground truth**

In determining the ground truth, we used the same approach as we did for the UMN dataset. At the beginning of all of the GTU clips, people acts normal. At some point of the clips, they try to evacuate the area immediately. However, some clips include only local abnormal events (LAE) where single or couple of people among the crowd is running. In GAE case, we select the initial frame of abnormal events when half of the people in the crowd starts running.

In Table 4.4, GTU dataset ground truth information is given. Each clip starts with normal situation. *NF start* denotes the initial frame number of normal events whereas *AF start* denotes the initial frame number of abnormal events.

Table 4.4: Ground truth for GTU Dataset.

GTU Dataset	NF start	NF stop-AF start	AF stop
Scene:1 Clip:1	1	278	480
Scene:1 Clip:2	1	396	491
Scene:1 Clip:3	1	322	484
Scene:1 Clip:4	1	132	218
Scene:1 Clip:5	1	NA	NA
Scene:1 Clip:6	1	1004	1120
Scene:1 Clip:7	1	1097	1193
Scene:1 Clip:8	1	NA	NA
Scene:1 Clip:9	1	254	335
Scene:1 Clip:10	1	391	474
Scene:1 Clip:11	1	1031	1130
Scene:1 Clip:12	1	1099	1213
Scene:2 Clip:13	1	1082	1215
Scene:2 Clip:14	1	760	953
Scene:2 Clip:15	1	602	703
Scene:2 Clip:16	1	476	556
Scene:2 Clip:17	1	384	473
Scene:2 Clip:18	1	1522	1614
Scene:2 Clip:19	1	311	395
Scene:2 Clip:20	1	1775	1857

4.2. GTU Dataset Performance Results

GTU dataset contains between 15 and 20 people, 20 video clips and two different scenes. Since we measured the R_far, and R_close values of the scene (more details are given in Section 3.3.3 about scene normalization), we implemented the scene normalization before evaluating the SEV graphs for each clip. From Section 4.2.1 to 4.2.2, we present SEV graphs for some of the video clips and the performance results for each scene.

4.2.1. Scene-1 Analysis

First scene of the GTU dataset contains 11 video clips all of which are captured in a cloudy day. Unlike UMN dataset, in this scene we test our method in LAE where only one or two people act abnormal. Additionally, in clip-8 and 5, there isn't any abnormal event. We have measured the scene normalization parameters such that

- $R_{far} = 74.5\ m$
- $R_{close} = 19.84\ m$

As mentioned in Section 3.3.3. we create normalization matrix by using the parameters given above. We will discuss the effects of scene normalization on the performances by giving the performance results of some clips without implementing the normalization. It is important to note that our method exposes some weakness most of which are based on false motions such as camera shaking, crossing birds or moving cars. SEV graph of clip-3 is given in the Figure 4.4. At the beginning of the video there is a camera shaking situation which results high SEV in the graph. These kinds of increases could result false alarms in thresholding step of the algorithm. In order to get rid of these kinds of errors caused by undesired motions, OF classification is necessary to ignore OF in these regions.

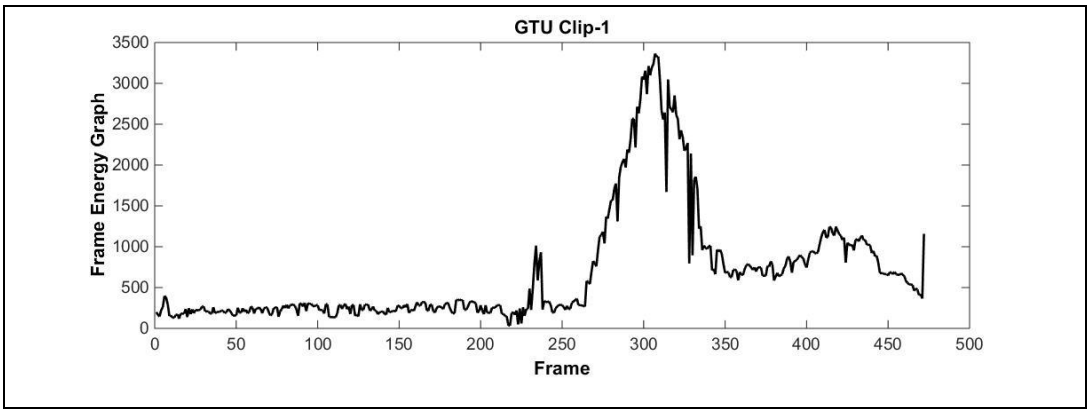

Figure 4.3: Clip-1 Scene Energy Graph.

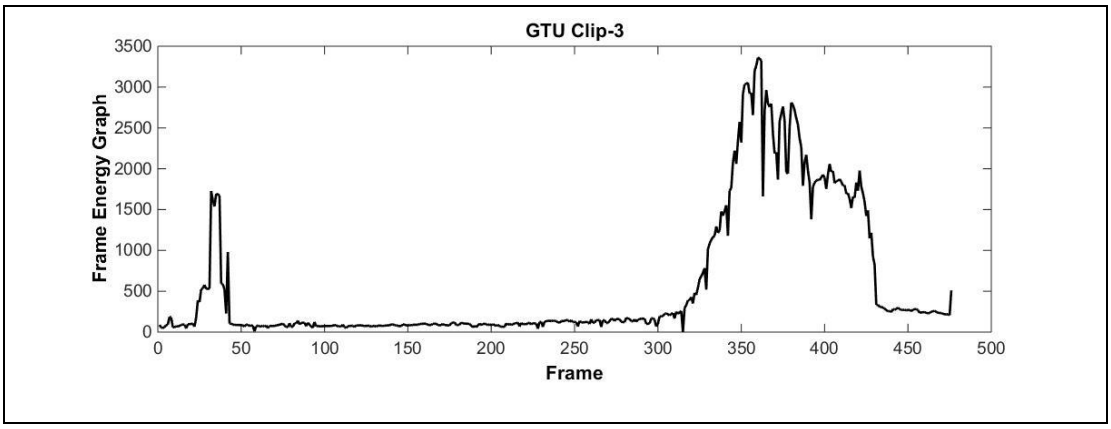

Figure 4.4: Clip-3 Scene Energy Graph.

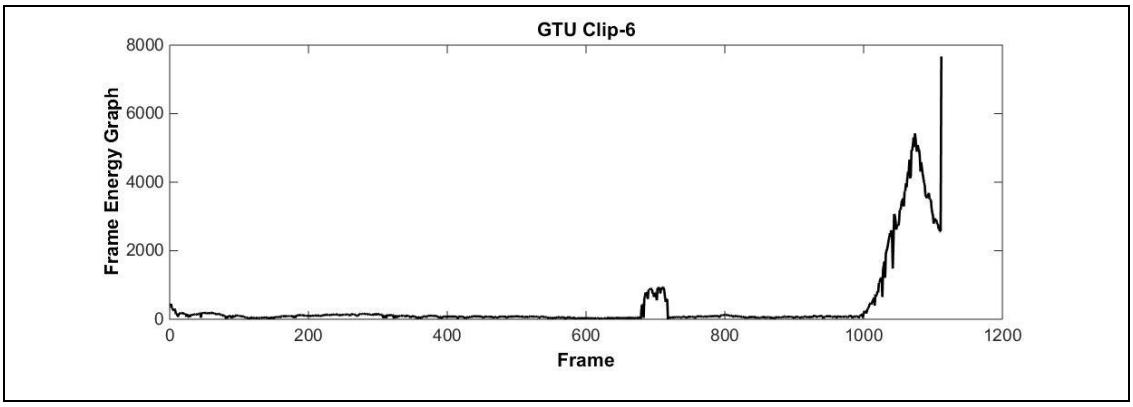

Figure 4.5: Clip-6 Scene Energy Graph.

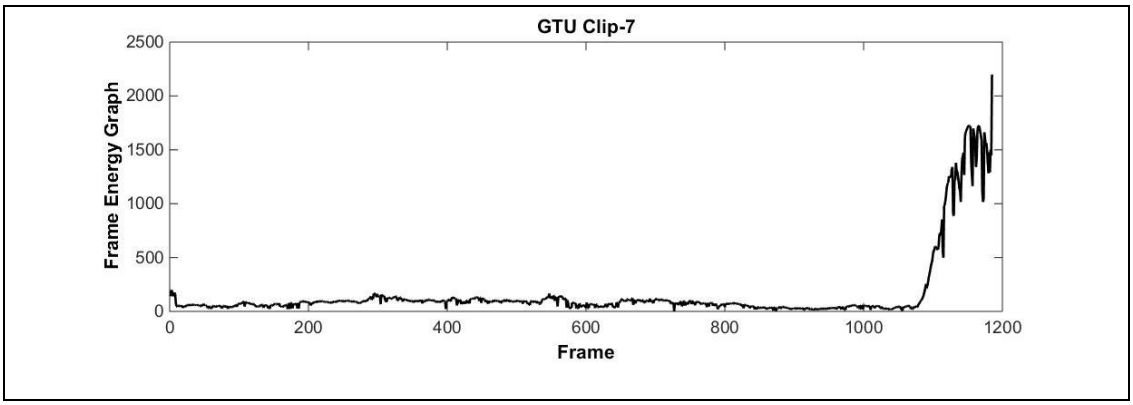

Figure 4.6: Clip-7 Scene Energy Graph.

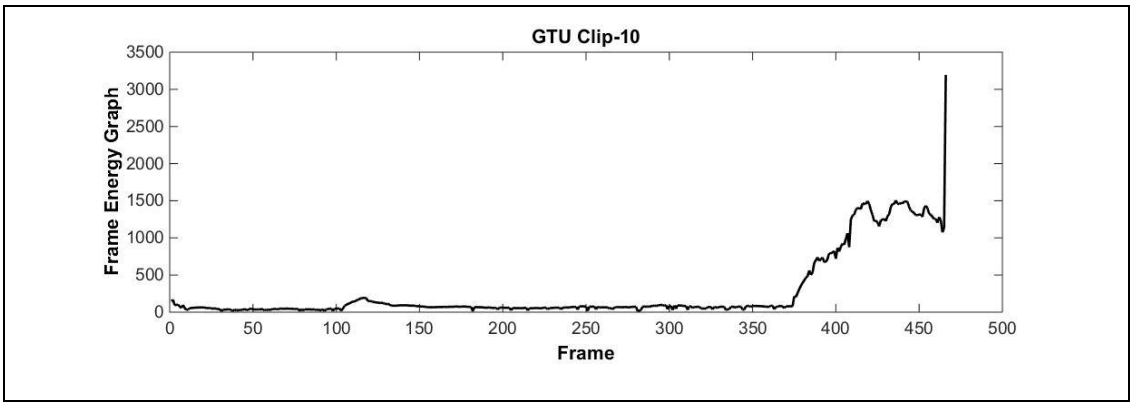

Figure 4.7: Clip-10 Scene Energy graph.

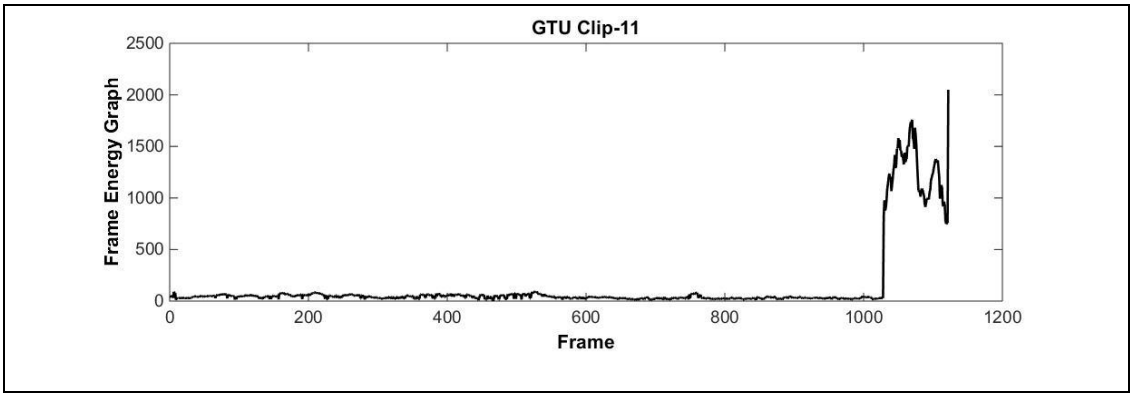

Figure 4.8: Clip-11 Scene Energy graph.

In Table 4.5, we have shared our GTU first scene performances table. To calculate each performance percentage, we leave the related clip out and extract a threshold value by using the rest of the clips. It can be noticed that clip-3 performance is lower than the 90 percent whereas the other clips performances. From

thresholding point of view, apart from the clip-1 performance both thresholding technique yield similar performances yet BF thresholding is slightly better than MN thresholding as far as performance percentages are concerned.

Table 4.5: GTU dataset Scene-1 Performance Results.

GTU DATASET Scene-1	Performance BFT	BFT Threshold	Performance MNT	MNT Threshold
Clip-1	%94.29	537	%85.41	685
Clip-2	%99.38	537	%98.34	763
Clip-3	%87.63	524	%87.42	655
Clip-4	%98.57	524	%98.57	736
Clip-5	%99.95	537	%99.95	779
Clip-6	%96.04	537	%96.31	721
Clip-7	%99.57	537	%98.56	764
Clip-8	%98.48	537	%99.15	733
Clip-9	%99.08	537	%97.25	745
Clip-10	%98.71	537	%98.71	747
Clip-11	%99.73	537	%99.73	717

4.2.2. Scene-2 Analysis

Second scene of the GTU dataset contains 9 video clips all of which are captured in a cloudy day. In this scene we have tested LAE cases as well. We have measured the scene normalization parameters such that:

- $R_{far} = 20.76\, m$
- $R_{close} = 11.46\, m$

It is important to note that, In clip-13 Figure 4.9, there is an abrupt change in SEV graph after frame number 200. It happened due to a crossing bird in front of the camera. Besides that, in clip-14 Figure 4.10, as people leave the area SEV of the corresponding frames drops to the normal levels.

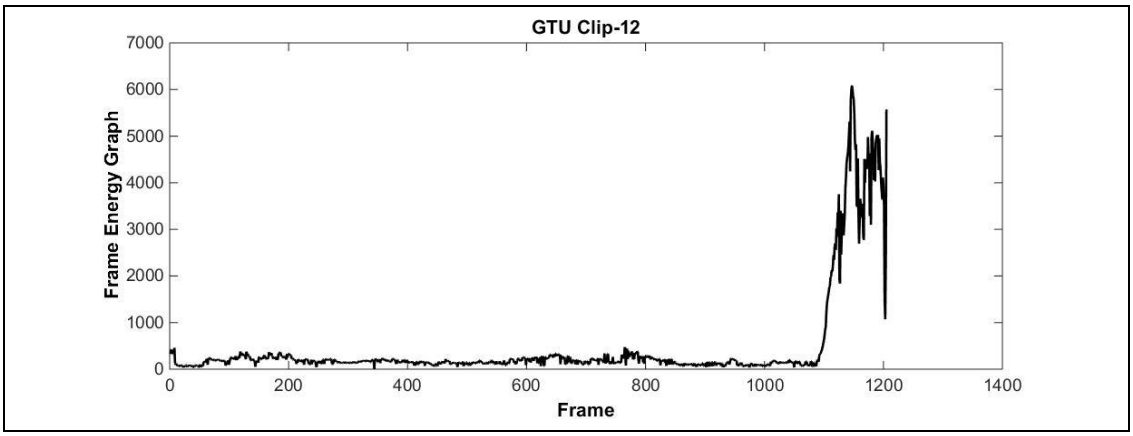

Figure 4.9: Clip-12 Scene Energy Graph.

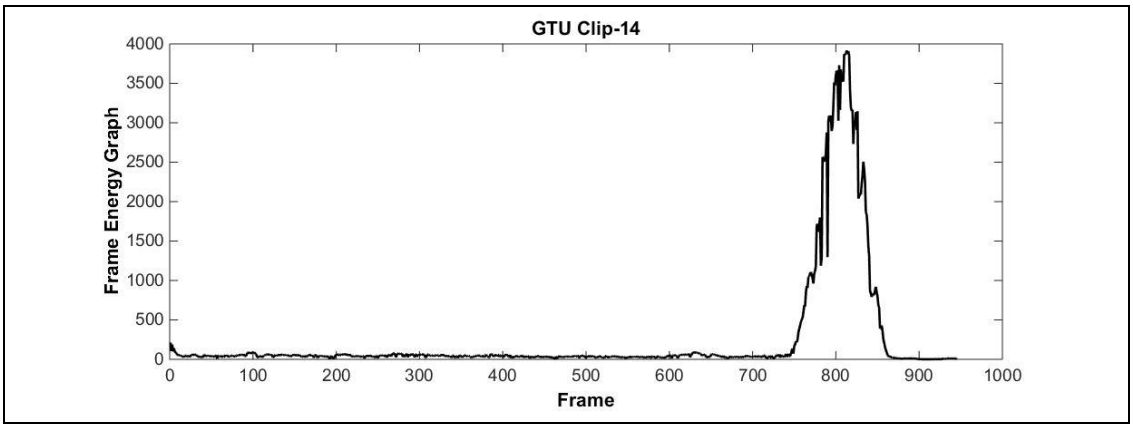

Figure 4.10: Clip-14 Scene Energy Graph.

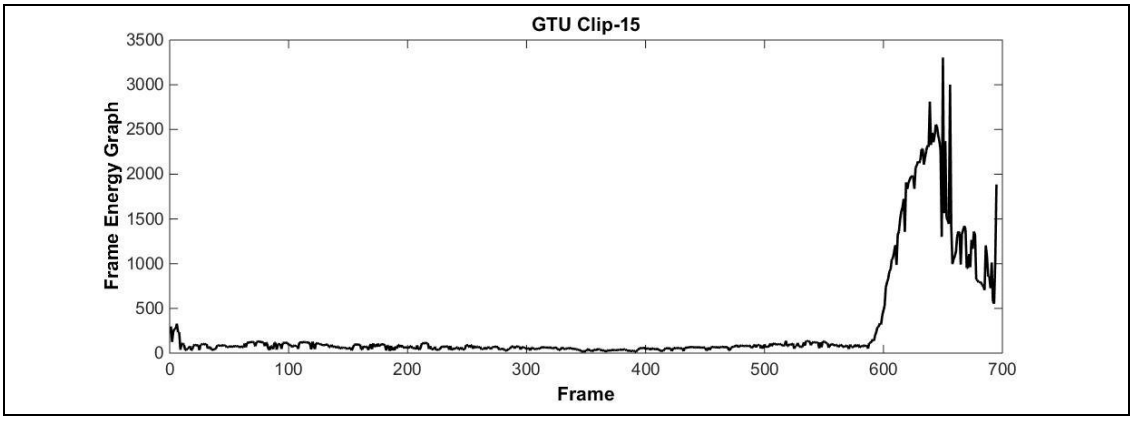

Figure 4.11: Clip-15 Scene Energy Graph.

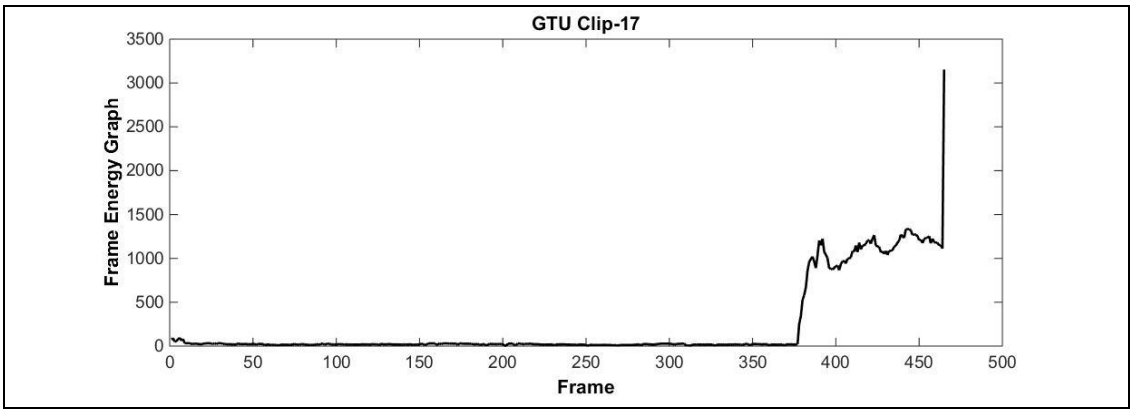

Figure 4.12: Clip-17 Scene Energy Graph.

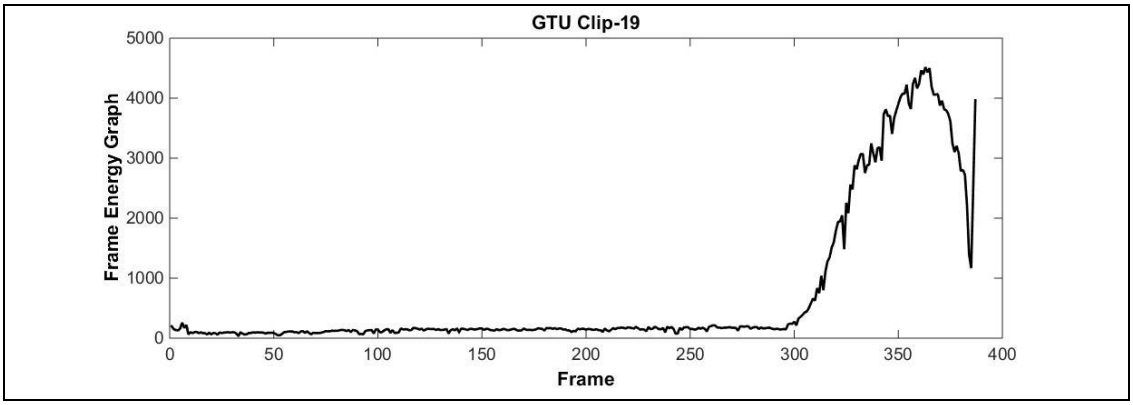

Figure 4.13: Clip-19 Scene Energy Graph.

In Table 4.6, we demonstrate scene-2 performances of GTU dataset. To obtain these performance values, we leave test clip out and the rest of the clips are used to evaluate a threshold value by using BF and MN thresholding methods. It can be seen that BF thresholding yields more accurate results than the MN thresholding. Only Clip-13 performance is lower than %95 in BF thresholding whereas clip-13 and clip-17 performances are lower than %95. The reason why clip-13 performance is noticeably lower than the other clips is that, crossing bird leads incorrect SEV values in a normal part of the ground truth.

Table 4.6 GTU dataset Scene-2 Performance Results.

GTU DATASET Scene-2	Performance BFT	Threshold BFT	Performance MNT	Threshold MNT
Clip-12	%99.91	671	%99.4	1230
Clip-13	%94.29	671	%94.9	971
Clip-14	%99.94	671	%99.8	850
Clip-15	%99.56	671	%94.7	1211
Clip-16	%99.81	671	%97.8	1228
Clip-17	%99.57	501	%88.2	1186
Clip-18	%99.75	501	%99.5	1235
Clip-19	%99.74	530	%98.5	1200
Clip-20	%99.45	530	%98.3	1224
Overall Performance/ Optimum T	**%99.08**	**671**	**%97**	**1149**

4.3. UMN Dataset Performance Results

UMN dataset contains 15-20 people for each video clips and three different scenes. Since the distance information is not given, the scene normalization step is not implemented. From 4.3.1 to 4.3.3, we presented SEV graphs for each video clips and the table of performance results for each scene.

4.3.1. Scene-1 Analysis

Video clips that belong to Scene-1 of The UMN dataset are captured in a sunny day where the background of the scene is mostly green and White textureless area.

At the beginning of the clips people act normal which can also be seen from corresponding SEV graphs where the extracted features take value somewhere around 0 to 30. After frame 480 for the first clip and 420 for the second clip there is an abrupt rise in SEV graph due to the initial running action of the people on the scene. Since we extract every SEV value from the grid that has maximum accumulated influence value, it is hard to make a comment about the pattern of the SEV graphs. However, it can be considered that, without the scene normalization, more people run/walk closer to the camera, greater its SEV value will be because amplitude of the OF of the pixels within the grids as well as the moving object size. will be greater.

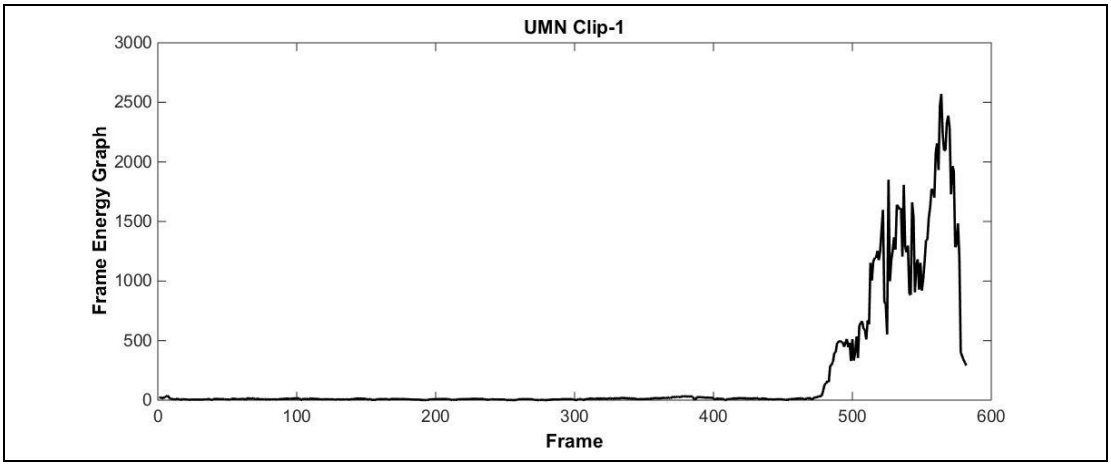

Figure 4.14 Clip-1 Scene Energy Graph.

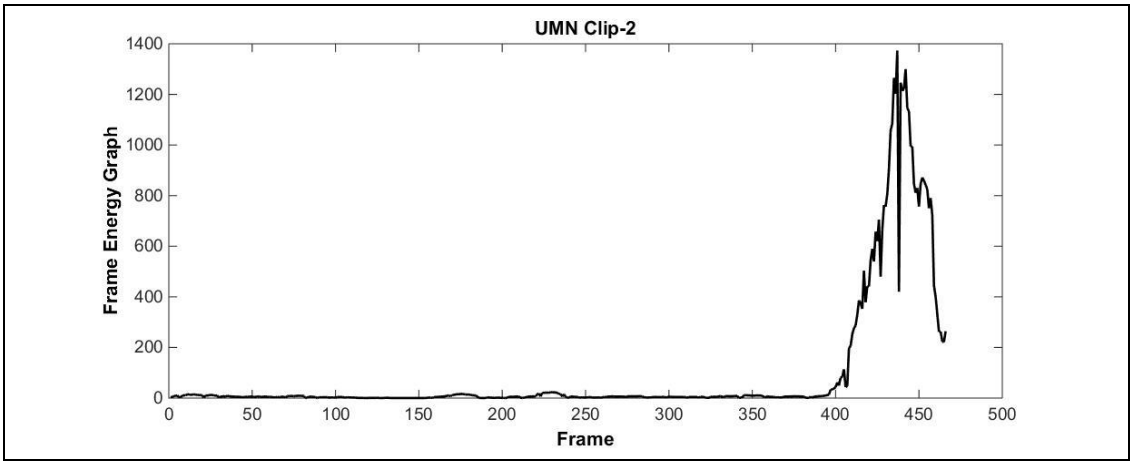

Figure 4.15: Clip-2 Scene Energy Graph.

In Table 4.7, we have compared our methods performance to other state of art methods in the literature. It can be seen that, BF thresholding yields approximately %1 better results than the MN thresholding. Since there are only two clips belonging to scene-1 we use one clip SEV data in order to calculate the other clip's threshold. Since camera angle and the height vary between scenes to scene we have to determine different threshold value for each scene. After averaging the obtained threshold values, one threshold value is found for the corresponding scene.

Table 4.7: UMN dataset Scene-1 Performance Results.

UMN Dataset Scene-1	Clip-1	Clip-2	T	
Social Force [5]	%96		-	
Streakline Potentials[8]	%90		-	
Optical Flow [5]	%84		-	
SAFM [9]	%98.6		-	
SVM[7]	%97.28		-	
Du [9]	%99.2		-	
Lee [6]	%99		-	
Wang [14]	%84.01	%83.36	-	
PROPOSED(BFT)	**%99.82**	**%98.07**	**113.5**	
PROPOSED(MNT)	**%99.31**	**%97.43**	**193**	**66**

4.3.2. Scene-2 Analysis

Second scene has some distinct properties in terms of the illumination than the other scenes in both UMN and GTU dataset. In low light condition individuals can not be distinguished from the background easily which effects optical flow accuracy. Since we use Brox' high performance OF [17] which yields more accurate results than traditional Horn-Schunk OF, second scene results is better than some similar methods in the literature([6], [14]).

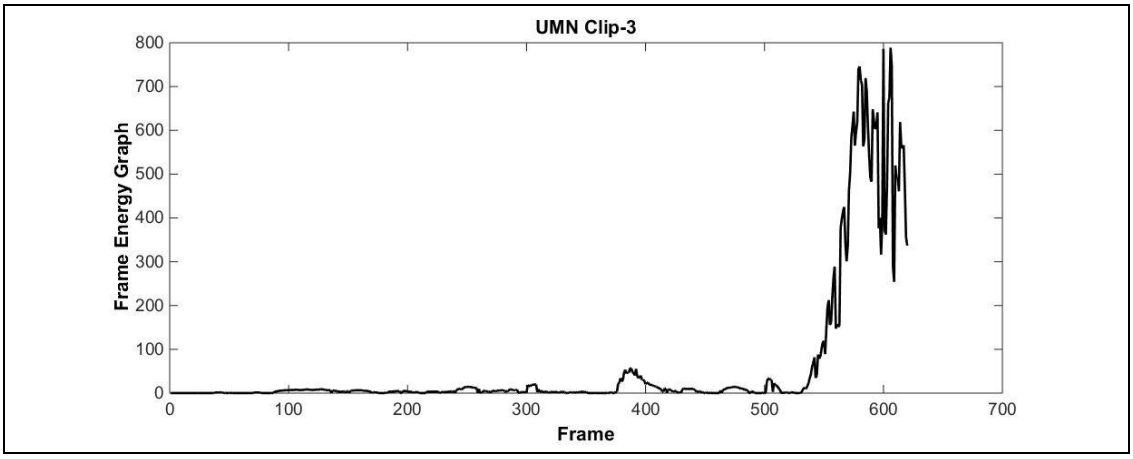

Figure 4.16: Clip-3 Scene Energy Graph.

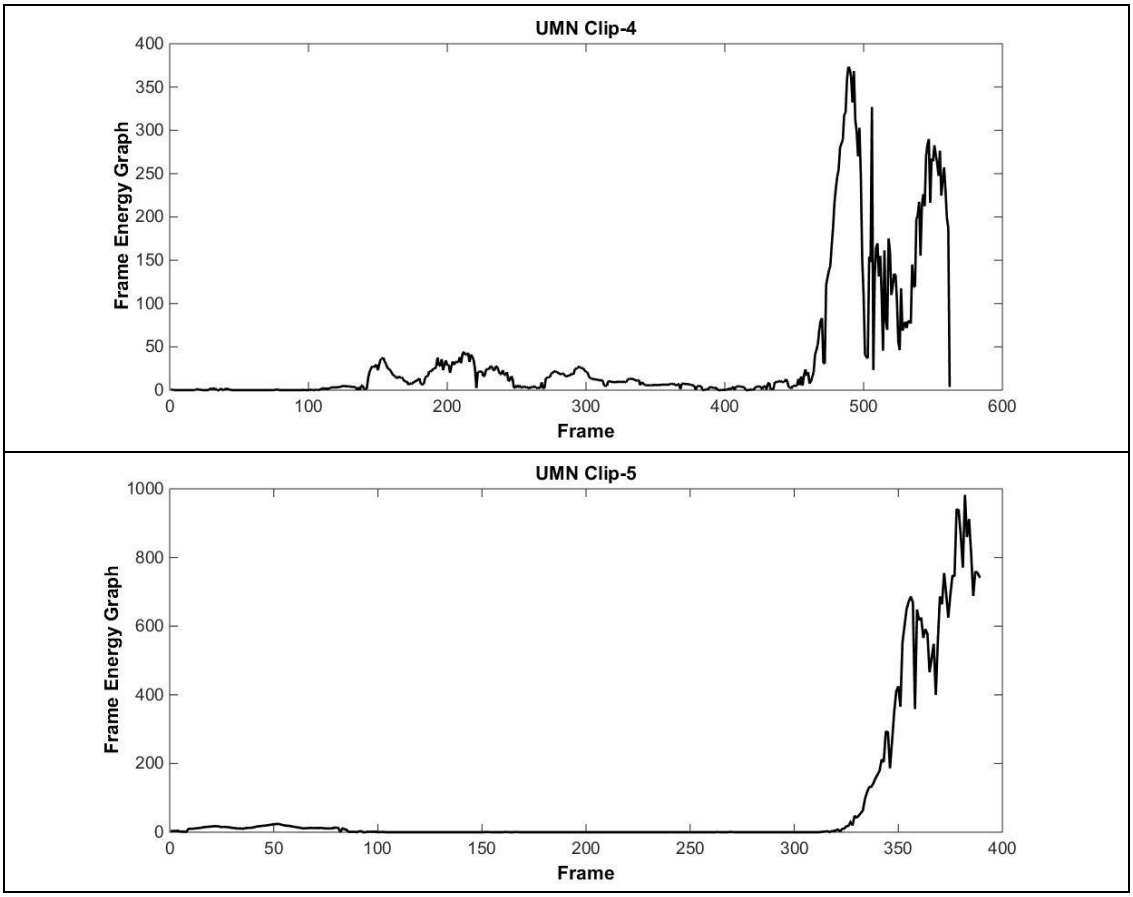

Figure 4.17: Clip-4(top) and Clip-5(bottom) Scene Energy Graph.

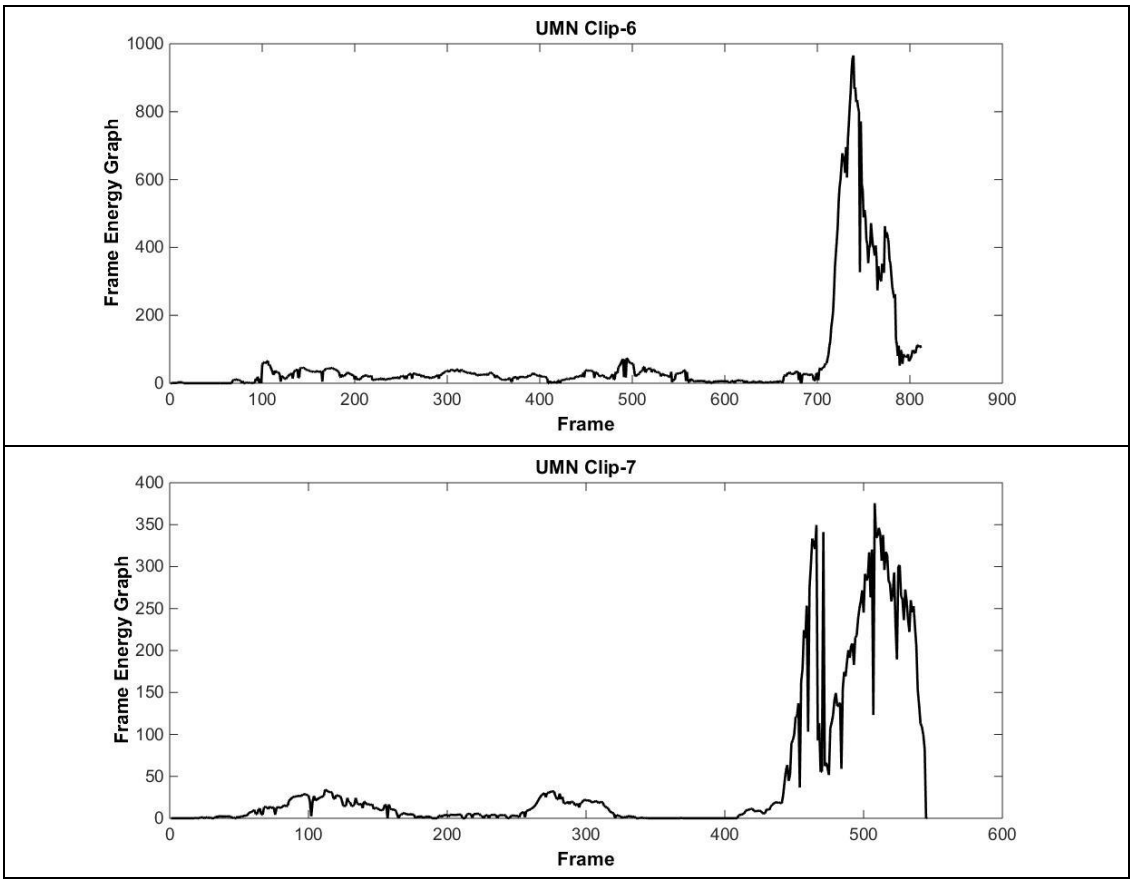

Figure 4.18 Clip-6(top) and Clip-7(bottom) Scene Energy Graph.

The second scene consists of five clips with different escape scenarios. In the Table 4.8, we have given our performance results of each clips as well as the other methods in the literature. It can be noticed that comparing the other scenes, Lee [6] influence matrix approach yields worse performance results in scene-2 which is directly related to the fact that traditional Horn-Schunk optical flow fails to produce decent flow due to the reasons we have mentioned at the beginning of Section 4.2.2. To obtain these values, we leave test clip out and from the rest of the clips we evaluate a threshold value by using BF and MN thresholding methods.

Table 4.8: UMN dataset Scene-2 Performance Results.

UMN Dataset Scene-2	Clip-3	Clip-4	Clip-5	Clip-6	Clip-7	Overall Performance/ Optimum T
SF [5]	%96					-
Streakline Potentials [8]	%90					-
Optical Flow [5]	%84					-
SAFM [7]	%98.6					-
SVM[7]	%97.28					-
Du [9]	%97.2					-
Lee [6]	%85					-
Wang [14]	%83.79	%83.63	%80.69	-	-	-
PROPOSED (BFT)	**%99.3**	**%97.5**	**%99.2**	**%98.5**	**%96.8**	**68.1**
PROPOSED (MNT)	**%99.5**	**%95.5**	**%99.2**	**%97.0**	**%96.7**	**72.84**

4.3.3. Scene-3 Analysis

Scene-3 is similar to Scene-1 of the same dataset where the clips were captured in a day light condition. One difference is the scene is closer to the camera comparing the other scenes. This distance differences can also be understood from the scene-3 SEV graphs given in Figure 4.18 and Figure 4.19. As people start running extracted SEV of frames are greater than the other scenes SEVs.

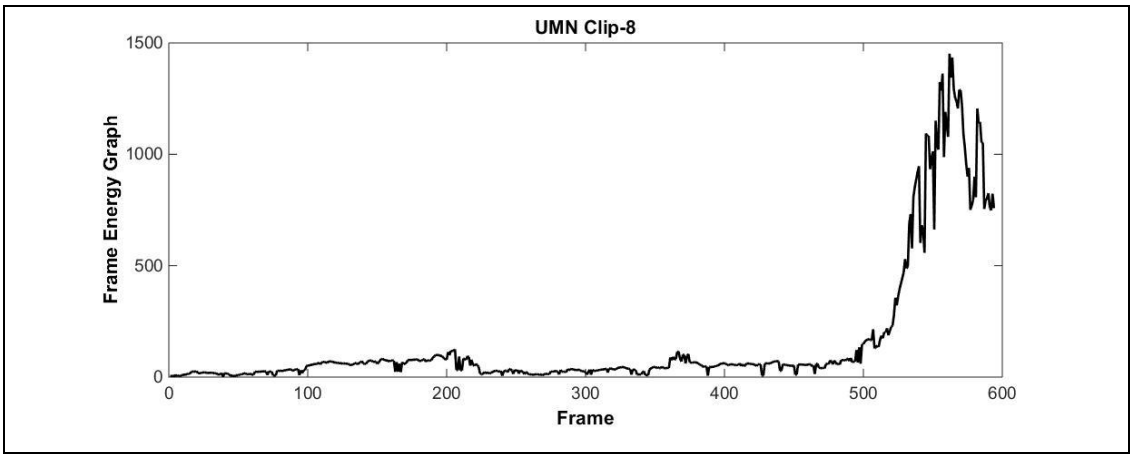

Figure 4.19: Clip-8 Scene Energy Graph.

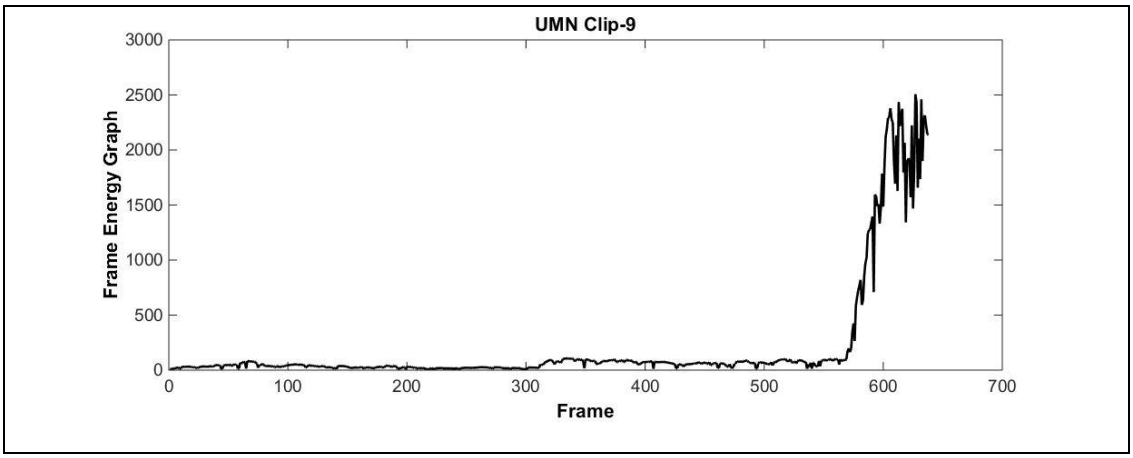

Figure 4.20: Clip-9 Scene Energy Graph.

In Table 4.9, we present the performance of the proposed algorithm and compare it with other state of art methods in the literature. As the number test clips is limited, we use leave-one-out methodology to determine each clip's threshold, i.e. we leave the interested clip out and determine a threshold based on the methods mentioned in Section 3.4 from the rest of the clips. Although each method produces a threshold different from each other, from performance point of view, it turns out that the performance value for these different thresholds are quite close. It happens due to the fact that SEV of scene-3 takes large values from 0 to 2500 and the thresholds are very close to each other compared to this large dynamic range.

Table 4.9: UMN dataset Scene-3 Performance Results.

UMN Dataset Scene-1	Clip-1	Clip-2	T
SF [5]	%96		-
Streakline Potentials [7]	%90		-
Optical Flow [5]	%84		-
SAFM [7]	%98.6		-
SVM[7]	%97.28		-
Du [9]	%97.8		-
Lee [6]	%98		-
Wang [14]	%86.88	%87.29	-
PROPOSED(BFT)	**%98.82**	**%99.06**	**401**
PROPOSED(MNT)	**%99.66**	**%98.74**	**261**

4.4. Effect of Scene Normalization

As it is mentioned in the Section 3.3.3, the distance between camera and the scene changes the magnitude of the optical flow vectors which directly effects the extracted feature values in order to classify video frames. This issue can raise false positives when the individual runs further away from the camera. To balance the extracted vectors considering their distance to the camera scene normalization process is implemented to GTU dataset. To explain the benefit of scene normalization we have extracted a SEV graphs of a person running around the camera scene in the Figure 4.21. The graph on top shows the SEV without the normalization, From frame 1 to frame 120 the person running closer to camera whereas after frame 350, the same person running with the same speed. It can be seen that normalization increases the feature values extracted furthest part of the scene thus, it could decrease the miss detection when the abnormal event occurs far away from the camera.

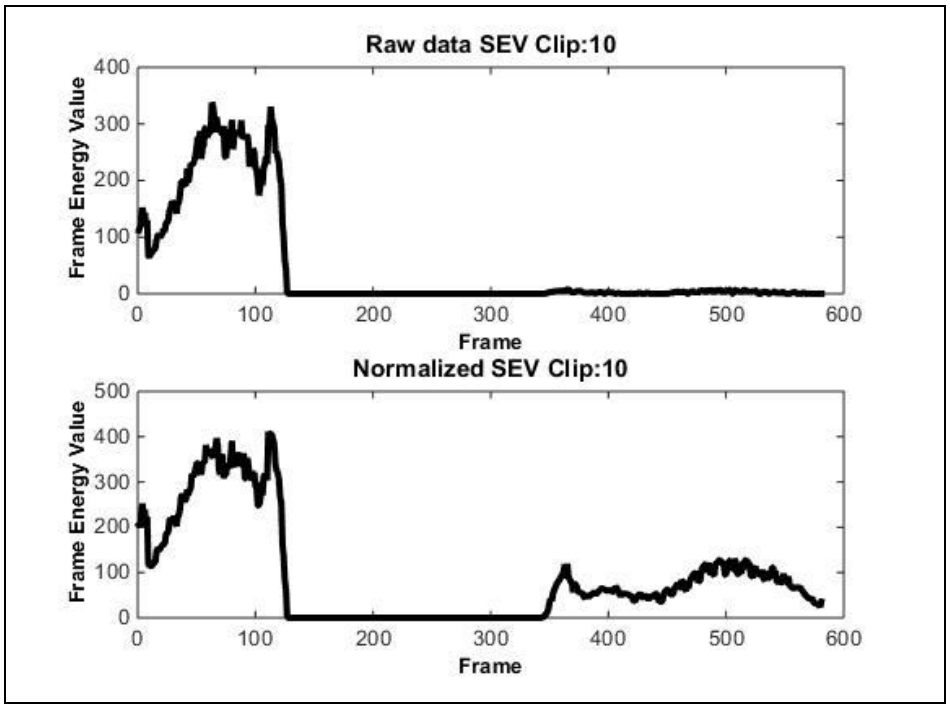

Figure 4.21: Comparison between Raw data and Normalized data.

4.5. System Properties and Calculation Time

We have given the average calculation times of our algorithm steps. It can be seen that high performance OF per frame requires most of the time whereas feature extraction per frame step requires less than half a second.

Table 4.10: Calculation time Table.

Process	Duration
High Performance OF	1.64 second / frame
Scene Energy Value Extraction	0.3 second / frame

System Properties;

- Processor: Intel(R) Core(TM) i7-2600K CPU @ 3.40GHz, 3401 Mhz, 4 Core
- Ram: 8GB DDR 3

5. CONCLUSION

In this thesis, we have covered the abnormal human behavior detection problem. Some famous both holistic and non-holistic approaches have been mentioned in the Section two. One of the main contributions of this thesis is that, unlike the other Works in the literature, we have used high performance optical flow algorithm in our proposed method. This accurate flow field helps to represent motion of the scene greatly and reduces the false alarms due to the noise which emerges from traditional optical flow algorithms. Besides, using novel grid based influence map approach as well as scene normalization, we have proposed an improved version of influence matrix approach that Lee[6] proposed. Using these influence maps, we extract single feature value to represent the level of abnormality of the scene. Using two different thresholding methods, we have tested our algorithm in two different dataset consisting of 29 video clips total. Overall performance results are promising such that, performance accuracy of 28 clips out of 29 are above %95. This work has also been presented in SIU2016 [28]. Despite these performance results, our algorithm still requires improvements in order to operate in real time. First issue of high performance optical flow algorithm is that, it takes longer than traditional optical flow to yield flow field. Other issue that should be taken care of is that, our method is not robust to objects moving in the environment which means that, any object (vehicle, bird or etc.) that moves fast in the scene will lead false error. In order to overcome this issue, object classification should be taken into account before calculating the influence map.

REFERENCES

[1] Brox T., Bregler C., Malik J., (2009), “Large displacement optical flow,” , IEEE Conference on Computer Vision and Pattern Recognition, 41–48, Miami, USA 20-25 June.

[2] Brostow G. J., Cipolla R., (2006), “Unsupervised Bayesian Detection of Independent Motion in Crowds,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1, 594–601, New York, USA, 17-22 June.

[3] Rosten E., Porter R., Drummond T., (2010), “Faster and Better: A Machine Learning Approach to Corner Detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 32 (1), 105–119.

[4] Basharat A., Gritai A., Shah M., (2008) “Learning object motion patterns for anomaly detection and improved object detection,”, IEEE Conference on Computer Vision and Pattern Recognition, 1–8, Alaska, USA, 23-28 June.

[5] Mehran R., Oyama A., Shah M., (2009), “Abnormal crowd behavior detection using social force model,”, IEEE Conference on Computer Vision and Pattern Recognition, 935–942, Miami, USA 20-25 June.

[6] Lee Dong-Gyu, Suk H.-I., Lee S.-W., (2013), “Crowd Behavior Representation Using Motion Influence Matrix for Anomaly Detection,” 2nd IAPR Asian Conference on Pattern Recognition, 110–114, Okinawa, Japan, 5-8 November.

[7] Zhang Y., Qin L., Yao H., Huang Q., (2012), “Abnormal crowd behavior detection based on social attribute-aware force model.”, 19th IEEE International Conference on Image Processing (ICIP), 2689–2692, Orlando, USA, 30 September-3 October.

[8] Zhao J., Xu Y., Yang X., Yan Q., (2011), “Crowd instability analysis using velocity-field based social force model.”, IEEE Visual Communications and Image Processing (VCIP), 1–4, Tainan City, Taiwan, 6-9 November.

[9] Du D., Qi H., Huang Q., Zeng W., Zhang C., (2013), “Abnormal event detection in crowded scenes based on Structural Multi-scale Motion Interrelated Patterns.”, IEEE International Conference on Multimedia and Expo, 1–6, California, USA 15-19 July.

[10] Kanade T., (1981), “Aniterative image registration technique with an application to stereo vision.”, Proceedings of Seventh International Joint Conference of Artificial Intelligence, 2, 674– 679.

[11] Kratz L., Nishino K., (2009), “Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models.”, IEEE Conference on Computer Vision and Pattern Recognition, 1446–1453, 41–48, Miami, USA 20-25 June.

[12] Kaltsa V., Briassouli A., Kompatsiaris I. Strintzis M. G., (2012), “Timely, robust crowd event characterization.”, 19th IEEE International Conference on Image Processing (ICIP), 2697–2700, Orlando, USA, 30 September-3 October.

[13] M. K., Skonieczny L., (2005), “Faster clustering with DB- SCAN,” Intelligent Information Process and Web Mining, 31, 605–614.

[14] Wang S., Miao Z., (2010), “Anomaly detection in crowd scene,” IEEE 10th International Conference on Signal Processing (ICSP), 1220–1223, Beijing, China, 24-28 October.

[15] B. H., Schunck B., (1981), “Determining optical flow,” Artificial Intelligence, (17), 185–213.

[16] Black M. J., Anandan P., (1993), “A framework for the robust estimation of optical flow,” in , Fourth International Conference on Computer Vision. Proceedings, 231–236, Berlin, Germany, 11-14 May.

[17] Brox T., (2004), “High Accuracy Optical Flow Estimation Based on a Theory for Warping,” presented at the 8th European Conference on Computer Vision, Springer-Verlag Berlin Heidelberg, (4), 25–36, Amsterdam, Netherlands, 8-16 October.

[18] Bruhn A., Weickert J., Feddern C., Kohlberger T., Schnorr C., (2005), “Variational optical flow computation in real time,” IEEE Transactions Image Processing., 14 (5), 608–615.

[19] Barron J. L., Fleet D. J., Beauchemin S. S., Burkitt T. A., (1992), “Performance of optical flow techniques.”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Proceedings CVPR ’92, 236–242, Santa Margherita Ligure, Italy, 19-22 May.

[20] Ju S. X., Black M. J., Jepson A. D., (1996), “Skin and bones: Multi-layer, locally affine, optical flow and regularization with transparency,” in Computer Vision and Pattern Recognition Proceedings CVPR’96, IEEE Computer Society Conference on, 307–314, San Francisco, USA, 18-20 June.

[21] Bab-Hadiashar A., Suter D., (1997), “Optic flow calculation using robust statistics,” in Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, 988–993, San Juan, USA, 17-19 June.

[22] Lai S.-H., Vemuri B. C., (1998) “Reliable and efficient computation of optical flow,” Int. J. Comput. Vis, 29 (2), 87–105.

[23] Alvarez J. W. L., (1998), “Reliable estimation of dense optical flow fields with large displacements,” International Journal of Computer Vision. 39 (1), 41–56.

[24] Bruhn A., Weickert J., Schnörr C., (2005), “Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods,” International Journal of Computer Vision., 61 (3), 211–231.

[25] Mémin E., Pérez P., (1998), “Dense estimation and object-based segmentation of the optical flow with robust techniques,” IEEE Trans. Image Process., 7, 703–719.

[26] Farneback G., (2001), "Very high accuracy velocity estimation using orientation tensors, parametric motion, and simultaneous segmentation of the motion field.", In Proc. Eighth International Conference on Computer Vision, 1, 171–177, Vancouver, BC, 7-14 July.

[27] Y.-T. Wu, Kanade T., Cohn J., Li C.C., (1998), “Optical flow estimation using wavelet motion model,” in Sixth International Conference on Computer Vision, 992–998, Bombay, India, 4-7 January.

[28] Tokta A., Hocaoğlu A. K., (2016), “Abnormal crowd behavior detection in video systems,”, 24th Signal Processing and Communication Application Conference (SIU), 697–700, Zonguldak, Turkey, 16-19 May.

BIOGRAPHY

Aybars Tokta was born in Istanbul/Turkey in 1991. He completed his high school education in Kartal Doğa collage in 2009 and graduated from Gebze Technical University (GTU) Graduate School of Natural and Applied Science, Electronics Engineering in 2014 with Summa Cum Laude. He has been working as a research assistant in GTU since 2015. His current field of study is Image, Video processing and Computer Vision. Along with academic efforts, he is giving free lectures on his youtube channel “Toktaakademi”.