Crowd detection and counting using a static and dynamic platform: state of the art

Chaudhry, Huma, Rahim, MSM, Saba, T and Rehman, A (2019) Crowd detection and counting using a static and dynamic platform: state of the art. International Journal of Computational Vision and Robotics, 9 (3). pp. 228-259. ISSN 1752-9131

A review on crowd detection and counting using a static and dynamic platform

ABSTRACT

Automated object detection and crowd density estimation are popular and important topics in visual surveillance research area. The last decades witnessed many significant publications in this field and it has been and still is a challenging problem for automatic visual surveillance. The ever increase in research of the field of crowd dynamics and crowd motion necessitates a detailed and updated survey of different techniques and trends in this field. This paper presents a survey on crowd detection and crowd density estimation from moving platform and surveys the different methods employed for this purpose. This review category and delineates several detections and counting estimation methods that have been applied for the examination of scenes from static and moving platforms.

Keywords:

Crowd; Counting; Holistic and Local Motion Features; Estimation; Visual Surveillance; Moving Platform.

1. INTRODUCTION

Fill An ever-increasing population issue is coupled with the occurrence of crowds and situations of overcrowding. The main motivation behind this research effort has been driven by an increased attention by computer vision research community towards crowd control and management. This has become a crucial issue especially due to ever growing world's population, and this increase directly relates to concerns regarding the security and safety of the larger population [29]. This motivation is tied also with the excess available surveillance data which is assigned to the limited amount of manpower available to process it.

In several real-world problems, the issue of identifying the number of objects, specifically people, in images and videos arises for different reasons including crowd creation alarm, crowd management, emergency evacuation of the crowd, design, and analysis of buildings and spaces for crowd management for safety and security. In certain scenarios, obtaining the people location and/or count is of direct importance, such as in public rallies, marathons, public parks, or transportation hubs. Manually identifying creation and movement of the crowd round the clock, or manually counting of individuals in very dense crowds is an extremely laborious task hence, several automation methods based on computer vision techniques have been proposed. Table 1 shows some of the deadly death rates in huge gatherings and highlights the significance of methods to prevent such cases in future.

This survey is intended to highlight the computer vision research on crowd detection for giving parallel insight to

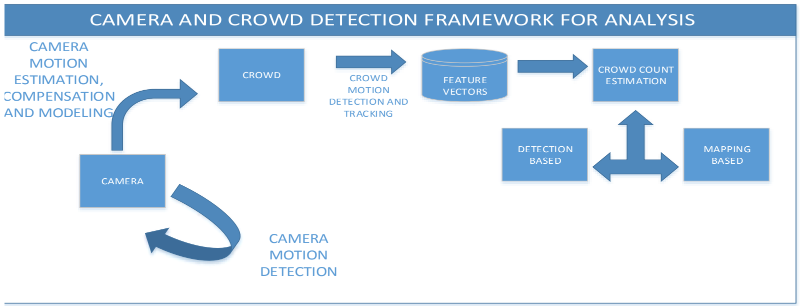


Figure 1. Crowd detection using moving platform, a framework

limitations of current research and how the recent work can be expanded to a more robust method of crowd detection from a moving platform. This work presents a survey of crowd analysis techniques based on computer vision while pointing out some open problems that could be further investigated and future directions. It should be noted that survey papers on crowd analysis have been proposed previously, such as, such as [40], [44] and [49]. However, Zhan et al.[40] focused on the pedestrian crowd in cluttered environments, while Junior et al. [44] covered survey on crowd behavior understanding and Saleh et al. [49] covered crowd detection and counting techniques that are applied to the sparse crowd only. However, none of the work surveyed literate on techniques and problems for crowd count and detection from moving platform. This gap is covered in this survey. Figure 1 illustrates the general framework for crowd detection techniques where the platform is a moving camera.

This work is organized as follows: Section 3 describes systems of Crowd density estimation and counting in different scenarios: Static camera based detections are briefly discussed in sub-section 3.1, Crowd video where the crowd is in motion is discussed in 3.2. Crowd detection strategies in the literature which are based on moving platform videos are discussed in sub-section 3.3. Section 4 shows Benchmark Datasets that are used in crowd density estimation and counting. Section 5 provides a general problem and future possibilities in Crowd density estimation research. Finally, the conclusion is presented in Section 6.

2. Crowd Dynamics

Crowd dynamics is a study of crowds; how and where they originate from, their behaviors, different motion states and from a more local perspective, how individuals in the crowd interact with each other, self-organize, and influence the overall crowd status [59, 60]. The taxonomy of crowd dynamics and platform dynamics has been illustrated in Figure 2.

Crowd dynamics is a non-trivial area of research. Crowds have been studies based on different observations which range from organized to unorganized, equilibrium to semi-steady state and static to violent. Hence crowd dynamics deals with many aspects of human nature. Motion state of the crowd, their inter and intra-spatial distance, and their size are some of the critical parameters which define the type of a crowd [62]. Crowd with varying spatial distance might show a region of interest, while motion pattern and speed might show emergency. If the crowd is assumed as a

Table 1: Example of some recent crowd disasters

Year	Place	Event	Casualties
2015 [1]	Saudi Arabia	Muslim Pilgrimage	1453
2015 [13]	Egypt	Egyptian Premier League	74
2013[24]	India	Hindu Festival	115

collection of entities, their interactions under different situations result in a change in their pattern of motion, spread, speed and other such factors. The individual entities can be influenced by the collective effect of other entities which will result in a collective behavior. A human crowd of sufficiently high-density exhibit features of collective behavior, such as lane formations [64], the emergence of clues or trails or bottlenecks near entrance or exits of a building [65]. To exemplify the complexity of crowd dynamics, an example of train formation is an example of motion patterns. Trails do not follow a strict straight line nor is it predesigned. Instead, it is a result of an optimization process of pedestrian flow for better efficiency, in which a number of individuals over time have contributed to the formation of the trail.

The standard on what measures the density of the crowd to be low, medium, high, very high, is yet to be established. The earliest of works suggested the five density levels in [67, 68]. Some other works suggested measuring the density per meter squared in an image [34]. However, of course in such case camera view, perspective distortion and occlusion pose challenges. A camera view of pedestrians, for instance, might classify a certain number of humans as the dense crowd. However, an aerial view, on the other hand, might not classify it as a dense crowd at all. Such confusions highlight the need for standardizing crowd dynamics for the different applications.

Understanding crowd dynamics is essential to manage the crowd. Some people are stationary in moving crowd, while among the static crowd there might be moving individuals, in a group or groups. Based on motion types of the crowd, crowded scenes can be divided into two categories [2]: structured and unstructured. The crowd which moves in a common direction while keeping motion coherence and main crowd behavior and motion direction does not vary frequently over time or space for a structured crowd. The crowd which is chaotic or contains varying, unpredictable crowd motion. Individuals move in different directions over time, and space while encompassing multiple crowd behaviors in the unstructured crowd [70]. In the field of visual image and video processing, the crowd dynamics has been studied in various ways to estimate better crowd count and behavior estimation. How these dynamics have a potential to be exploited is discussed in detail in the following sections.

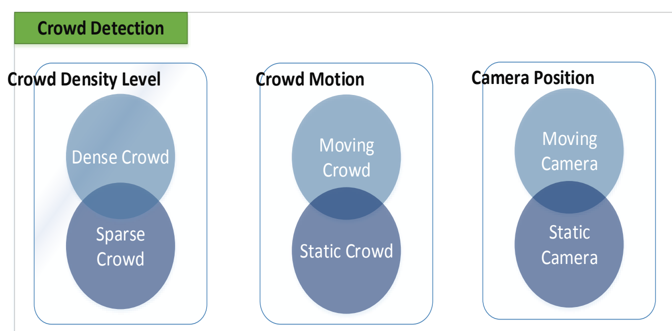


Figure 2. Taxonomy of Crowd Detection in research classification

3. DETECTION AND COUNTING METHODS FOR CROWDS

In this section, we analyze which features of a crowd's motion are used to detect and count crowds in various scenarios, such as static crowd and moving crowd. Also, in the case of camera motion, we review the recent literature which covers methods used and present technological limitations and challenges for crowd detection and counting in such scenarios.

3.1 Crowd Detection using Static Platform

Several studies on crowd detection methods have been surveyed in detail over the years [40, 44, 49]. Since this area has recently been surveyed, this paper will not cover crowd detection methods from the static platform. For a more recent survey on crowd detection and counting methods, the reader is redirected to [49]

The crowd can be in motion as well as quasi-static or static, at the same time. Most of the crowd detection methods assume a single state of motion when dealing with a crowd [73]. While this is not the case always. Some people are stationary in moving crowd, while among the static crowd there might be moving individuals, in a group or groups, as mentioned earlier. The analysis of moving crowd is done at global and local levels in the literature, where motion is the basis for crowd analysis. The analysis is done for several purposes: motion detection, event detection, behavior understanding, counting or segmentation of crowd. The global and local level of moving crowd's analysis can be termed as macroscopic and microscopic levels, respectively. At the macroscopic level, analysis targets the global

Figure 3. method used by [3] which tracks objects for a longer time and also maintains more tracks than the method in (b) by [21].

motions of a huge number of people, hence mapping-level detections and count estimations are used, also known as, regression-based estimation. Whereas for microscopic crowd analysis, direct detection based approach is used for detection and count estimation.

In the next sub-section, we will be exploring motion-based features for crowd detection at the local and global level, while also highlighting the ways of estimating the crowd count using the motion information.

3.1.1Microscopic Level Detections:

Local level information is essential to understand the origins of multiple crowd behaviors. At the microscopic level, the interaction of the crowd's elementary individuals can be studied and used to analyze the resulting crowd behavior.

MOTION MEASUREMENTS: Pixel-level features help to extract Microscopic level significant information in a scene. Densely extracted Flow-Based Features are one example of local spatiotemporal level features extracted from crowd motion based on local information from 2D patches or 3D cubes. Motion features derived from consecutive video frames is a type of pixel-level feature, known as Optical Flow [74]. The formula for optical flow on image I am given as

$$I(x, y, t) = I(x + dx, y + dy, t + dt)$$

which is approximated with Taylor series and derivating with dt results in describing optical flow in spatial image gradient form. It can be seen in the following equations

$$f_x u + f_y v + f_i = 0$$

where

$$f_x = \frac{\delta f}{\delta x};\ f_y = \frac{\delta f}{\delta y}$$

$$u = \frac{dx}{dt}, v = \frac{dy}{dt}$$

where u and v vectors represent a change in horizontal and vertical speeds, respectively.

The motion within each local area may be non-uniform and generated by any number of moving objects. Optical Flow works well for slow moving objects. Robust Local Optical Flow (RLOF) performs even better for sparse local information [3].

Several optical-flow like features analysis methods avoid tracking from the macroscopic level [6, 17, 32, 75, 76]. Such methods have found applications in addressing complex crowd flows in the scenes. Particle flow is one method for such analysis, such for motion segmentation [6]. It provides trajectories which relate a particles' position at its initial stage to its later stage. Optical flow is robust to camera and object motions [17, 45, 77-79]. R. Mehran [32] introduced the notion of streamline to get a more accurate motion field for crowd video scene analysis, referred to as streak flow. Streaklines encapsulates motion information of the flow for a period. Some extremely crowded scenes are less structural.

Motion and Keypoints: Optical flow method has been combined with several key point features for better moving object detection, such as Haris corner in [41, 47] . FAST key points used in [9], and SURF features used in [27], the results are shown in Figure 4, or tracks extracted by Kanade-Lucas-Tomasi (KLT) [80], to name a few.

Figure 4. (a) The results of trajectory clustering and independent detection motion for counting people on USC dataset. (a) Method proposed by G. J. Brostow [5]. (b) Method proposed by Rabaud and Belongie(2006). (c) Method proposed by Donatello Conte [22] (d) Method proposed by R. Liang [27]

Table 2: Some of the microscopic flow based features advantages and limitations

Feature	Advantage	Limitation
Particle Flow	Can track for longer using initial and current position, Time Averaged OF.	Time delay is significant Spatial changes are ignored
Streak Flow	Captures instant changes in the flow than particle flow,	Better crowd motions results in dynamically changing flow.
Optical Flow	Robust to Camera, Object Motion	At a microscopic level, for a dense crowd, can be incompetent
Motion Histogram	Only advanced MH methods are more suitable, such as • Multi-Scale Histogram of Optical Flow (MHOF), o motion and spatial information • Orientation Distribution Function (ODF) o Computationally effective	Computationally expensive, Motion alignment problems with limited quantized directions

Motion and Counting at Microscopic Level: At a local level, several methods have been proposed which use different measurements for counting. Local level counting is achieved using direct methods [81-83]. Also, linear relationship between detected blob size and group count has been used for motion-based counting [81, 84]. Kilambi, Ribnick [82] used detection based counting, where elliptical shape model was used for people counting. In [85], a detection based method is used, where a 3D model of the human body is mapped to the detected foreground regions using a Markov Chain Monte Carlo (MCMC) approach. MCMC performs a global optimization of the posterior probability across multiple frames (in order to exploit temporal coherence). Figure 3 shows counting methods using Keypoints and motion features.

The number of matched detections is the count of people. Hajer Fradi [3] used the relationship of higher density of trajectories with a higher crowd density. Additionally, some studies propose that number of clusters are proportional to the number of people. Donatello Conte [22] used motion vector clusters to estimate count using regression-based methods. This method of G. J. Brostow [86] seems to perform well, even with high crowd densities, However, the method can have problems in low-density conditions, where the motion of arms and legs is clearly visible, because of its rigid motion assumption

3.1.2 Macroscopic Level Detections

Global level information is essential to understand the overall crowd motion trend and crowd behaviors. At the macroscopic level, the holistic properties of the scene are modeled. Assuming that high-density crowd behaves like a dynamic system, many dynamical crowd models have been proposed, [6, 32, 87, 88], which is based on the concepts of motion field and dynamic potential borrowed from fluid dynamics community [89]. Fluid dynamics finds its way in crowd motion analysis as people do not always follow the laws of physics; they have a personal choice for motion direction and do not bound to laws of conservation of momentum hence stopping and starting at will [59]. In highly crowded surveillance scenes, sometimes it's better to be observed at the macroscopic level. Moving objects in the sensor range appear small or even unresolved and very few features can be detected and extracted from an individual object. Hence, getting the larger picture of crowd motion by analyzing their behavior and anomaly detection can be more advantageous.

Table 3: : Global level motion analysis features, their applications and limitations

Feature	Advantage	Application	Limitation
Trajectory	Simplistic model to find object motion in a scene	To find relative distance between objects, acceleration, or motion energy. For longer intervals	Poor performance in accurate object detection and finding complete trajectories [8] as crowd density and scene clutter increase.
Tracklet	More conservative and less likely to drift than long trajectories.	learning semantic regions and clustering trajectories [10-12] and connected into complete trajectories for tracking [14-16],	Better for Local Level motion analysis Terminate when ambiguities caused by occlusions or scene clutters arise
Optical Flow	Robust to Camera, Object Motion	Several	Spatial and Temporal properties of the flow cannot be captured

Commonly Used Features For Macroscopic Motion Analysis: Trajectories and tracklets are the most commonly used features to analyze crowd motion and activities at a global level [90]. As crowd motion is normally regular and continuous across the frames in a normal scene, these holistic level features can help in analyzing the crowd movement at large. Using trajectories of people and objects, different levels of information can be extracted, such as direction, acceleration, relative distance between objects, or motion energy . Tracklet is a fragment of a trajectory obtained by the tracker within a short period. These features, trajectory or tracklet, compute the individual object's tracks, as the basic features for motion representation. In [87], tracking is used to analyse crowd while participants of the are assumed to follow a similar motion pattern. The floor fields proposed in this work impose a constraints on pedestrian motion direction, which results in only one single direction at each spatial position in the video. M. Rodriguez [2] extended the approach in [87] to unstructured environments where the assumption of fixed direction motion does not hold. The results of their work are shown in Figure 5. Thereby, their method can handle different crowd behaviors originating from overlapping motion patterns, however the direction is quantized in 10 directions.

Motion Based Features for Counting at Macroscopic Level: At a global level, several methods have been proposed which use different measurements for counting. Indirect counting methods use measures such as amount of blob size

Table 4: Brief summary of motion pattern analysis techniques

References and Dataset	Motion Features	Static Feature	Scene	Density level	Motion Analysis Method
Shah [6], M. Hu [17] UCF	Particle Flow	-	S	H	Flow Field Model Segmentation
[3]UMN	Robust Local Optical Flow	FAST corner points	U	M	GMM segmentation
R. Mehran [32] UCF	Streak Flow	-	S	H,	Flow Field Model Segmentation
Wong [35],Liu [38] UCF / UCSD	Optical Flow	SIFT, KLT	S	H/M/L	Flow Field Model Segmentation
Radke [42] Cheriyadat	Optical Flow	Harris	S	L	Similarity Based Clustering
Y. Yang [45],I. Saleemi [51], L. Song [53] MIT Traffic	Optical Flow	-, -. -	S	L	Probability Model Clustering
B. Zhou [55] UCF / Zhou	Tracklet	KLT	S / U	H/M	Similarity Clustering
C. Wang [12] UCF	Tracklet	-	S / U	H/M	Similarity Clustering
B. Zhou [8]CUHK	Tracklet	KLT	S / U	M	Probability Model Clustering
P.-M. Jodoin [58] MIT Traffic / UCF / Jodoin	Motion Histogram	-	S / U	H/M/L	Similarity Based Clustering
W. Fu [61] UCF / QMUL	Motion Histogram	-	S / U	H/M/L	Probability Model Based Clustering

*H= High, M= Medium, L= Low *S=Structured U=Unstructured

[91], moving pixels [92], fractal dimension [93], texture features [94] or key points [95]. In [95], corner points are

found using a variant of the popular Harris corner detector [96], after which background corner points are divided from foreground corner points. Background and foreground are distinguished using an estimate of the motion vector based on block matching between adjacent frames: points below a threshold motion speed are not considered as interest points. As a final point, the number of people is estimated from the number of moving corner points assuming a direct proportionality relation. Keypoint based methods pose problems such as instability of the corner detector, the presence of occlusions, or perspective error. In case the counting methods fail due to broken trajectory problem, the Super Track method can be used which has been proposed by [17] to solve this issue.

Figure 5. Examples of detection and tracking results for different crowded scenes and true positives are shown in green. . M. Rodriguez [2]

Table 5: Suitable features based on scene characteristics

Scene	Characteristics	Features
Outdoor Scenes,	Captured often with wide field of view Low resolution for each target	Aim: to analyze the holistic crowd motion trends. The optical flow alike features are suitable. Can be combined with flow-based models. (FBM has priority in structured crowded scene analysis).
Indoor Scenes,	The resolution of a single target is high enough. The crowd density may not be so high.	Aim: to analyze the local (or global or both) motion trends. The object trajectories or tracklet-based features are more suitable. For analyzing activities or semantic regions, can be combined with agent-based models. More suitable for unstructured crowded scenes.
Both	When the field of view is not wide. crowd density is high with severe occlusion	Aim: to analyze dense crowd in limited field of view Various local spatio-temporal features could be considered Optical flow alike features or tracklets not suitable

3.1.3 *Combining Microscopic and Macroscopic Detections*:

Results from both levels, microscopic and macroscopic, could be jointly used. Table 4 shows the motion and static features used commonly. For instance, [2, 97, 148,149] employ the information of scene features, as illustrated in Table 5, and global level motion to improve tracking individuals in a crowd scene, while the microscopic information of individual movements can be used as basic unit in the holistic scene analysis, as shown in Figure 6. Marco Manfredi [98] dealt with the challenge of the multi-motion state of the crowd in a scene. Two camera based systems were used to address the problem; master (larger motion view), slave (local motion information from ROI).

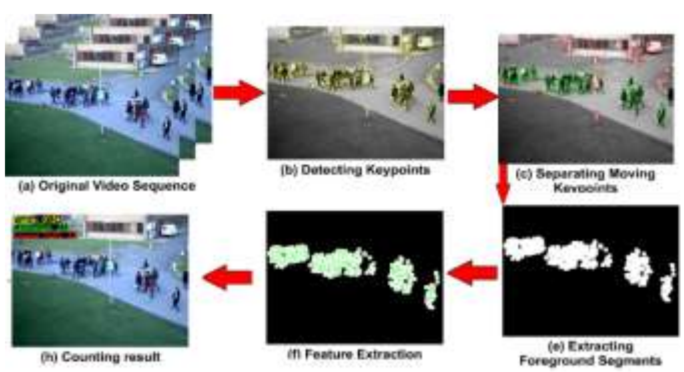


Figure 6. The most recent method for crowd counting using motion features proposed by [9]. This figure shows the detailed overview of the proposed solution. Count is estimated with foreground and FAST keypoints.

In heavily crowded scene: congestion level, the degree of occlusion, the size of field view of the camera, and the resolution information of a single target in the scene, the most suitable features based on the scenario is listed in the table below:

3.2 Crowd Detection from Moving Platform

The crowd can be captured using a still camera image, video or moving video. Assumptions made for video captured from stationary camera do not hold for a video captured from mobile devices, such as mobile phones or camera mounted on moving car, UAVs, or UGVs, as they raise several new challenges. An event captured from a moving camera notices several types of motions other than motion of objects in the scene [99, 100, 150-152]; the motion of the camera or changing scene geometry, or motion of platform carrying the mobile device. Moreover, each motion induces a different effect based on factors such as camera distance from captured event, the velocity of the capturing camera, and objects' velocity in a scene. This also results in a changing scale of the crowd. If an object under analysis is close to the camera, such as a vehicle shown in Figure 7, the motion can be captured locally in the frame. However, where human motion is concerned, due to the geometric constraints, for instance, the smaller region covered per frame, as shown in Figure 9, human motion detection from a moving camera can be more challenging.

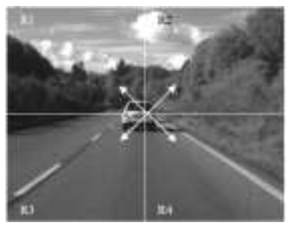


Figure 7. Region Division for Estimating Motion Vectors [23]

Pedestrian detection, from upfront camera view, as shown in Figure 8, wins an advantage of the geometrical view. The pedestrian size is large enough for motion analysis. However, when the pedestrian is far from the camera, motion features are not very distinguishing. Hence, some weak results can be seen in Figure 8.

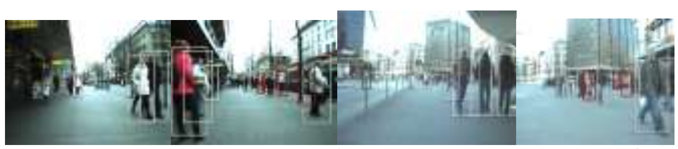


Figure 8. Experimental results obtained on 3 test sequences. Note the level of interaction between pedestrians (frequently overlapping bounding boxes). The images also show some typical false positives in red (trees, children's stroller, signs, mannequin).[25]

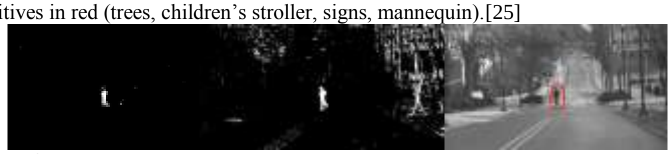


Figure 9. Experimental Detecting results of Bayesian decision method (a), which detects human accurately as the moving object in (c) .The threshold results show a reduced background noise (a) as compared to the methods which use geometric view constraints (b). [30]

As motion compensation lets camera motion effect to be reduced, it has often been used in application of moving camera where scene is planar or quasi planar, as shown in Figure 7. RANSAC is used to compute best estimate of motion trajectory subspace. Markov Random Fields have been applied in several works [101-103, 153,154] for better motion region detection in a scene captured from moving platform. In this section we discuss the motivation for crowd detection from moving platform and the research trends in this field. Table 6 illustrates some of the recent research for crowd detection from moving platform.

Pedestrian detection from a moving vehicle, which provides person view from an upfront angle, has been active since more than a decade [104-106]. However, work on crowd detection from an aerial view is more recent and is still evolving. Hence, to the best of our ability, we cover object detection from an aerial view and analysis methods that can be useful for dense crowd detection from an aerial platform.

Table 6: Some of recent research on crowd detection from moving platform

Author	View	Object Detection	Benchmark Dataset
Reisman, Mano [18]	Ground-based	Crowd	Reisman, Mano [18]
Ghosh [36]	Ground-based	People	UCF Sports Active
Hu, Chen [41], [47]	Ground-based	Pedestrian, Vehicle	Hu, Chen [41]
Wan Yanli [54]	Ground-based, PTZ	Person	Wan Yanli [54]
Ess, Leibe [25]	Ground-based	Pedestrian	Ess, Leibe [25]
Lin [63]	Ground-based	Person, Vehicle	Lin [63]
DeGol and Nam [7]	Aerial	People	VIRAT aerial data set
Perko, Schnabel [66]	Aerial	Crowd	Perko, Schnabel [66], [68]
Jong Taek, Chia-Chih [69]	Aerial	Person, Vehicle	Jong Taek, Chia-Chih [69]
[71]	Aerial, Ground-based	Car, Human	ViBe, PETS, [71]
[72]	Aerial, Ground-based	Car, Human	[72]

3.2.1Camera Mounted on Ground Based Vehicle

Research in this area, such as [157,158] targeted pedestrian detection from moving vehicle. Early method's common approach was two-step: combine stereo vision first, then apply some classification or result validation. Jalalian and Fathy [23] suggested that as the background is changing, frame differencing alone will not correctly detect the moving objects so they proposed Representative Motion Vectors. The representative motion vector of a region is the most repeated motion vector in that region. It can be visualized as HOG, only that it's a histogram of Motion Directions, extracted for each region in an image divided into four parts, as shown in Figure 7 earlier.

Some of the recent experimental results in this regard are shown in Figure 8 and Figure 9. It is noticeable that background and occlusion are some of the main challenges in such field of view. To find of pedestrians, in line with the hypothesis that the pedestrian size is proportional to camera distance, they use features such as size and parameters of the objects to use with skeleton, edges and signature as used in other related works [36, 54, 107, 108] of each region, for detecting of human parts, such as legs, for instance [159,160]. For tracking detected pedestrians, [41] used Kalman filter which defines the pedestrian region as a center of gravity of a moving object in a minimum bounding box. In a more recent work, a different background, foreground segmentation method was proposed, after motion compensation, done using Bayesian Decision Method and Belief Propagation method, which used motion feature vector.to segment people [30]. These methods successfully employ epipolar geometrical constraints for people detection, using information such as the angular difference between epipolar lines [72]. Epipolar geometry is used to find the relation of the points in two corresponding camera's view, Epipolar geometry is independent of the size, color, or shape of the object of interest. Another work by used epipolar geometry, [109] and proposed an extension to it, achieving better results. However, their assumption of individual pedestrian's absolute segmentation made the system prone to poor results for a more complicated scenario where occlusion or clutter was high.

3.2.2 Object Detection from a Moving Aerial View

The previously mentioned systems focus on pedestrian detection and not crowd detection which is usually captured from a far distance and uses motion detection in scenes captured by a camera onboard an aerial vehicle. In scenes captured from an aerial vehicle, a system faces several challenges, as an example is illustrated in Figure 10.

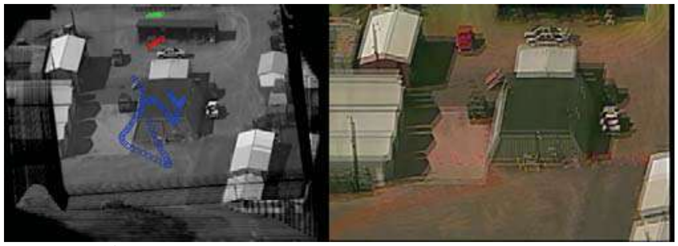


Figure 10. Common challenges in scene analysis from aerial view. The camera motion is shown in blue in (a), undergoing jitter and undesirable dynamic motions. Distinguishing slow object motion, depicted in red and green, is difficult from the irregular camera motion. This random camera motion, (b), induces motion blur and double image effects [7]

Traditional methods for motion detection from video captured by a moving camera include background subtraction with the help of color and motion features [110-112]. Other works, such as [47] use keypoint features to across the frames to segment foreground and background. However, the center of a PTZ (Pan-Tilt-Zoom) camera (or a camera on a defined rotation axis at a platform) is still fixed, unlike that of a moving camera, which has more dynamic, uncontrolled motion. Therefore, background subtraction fails for crowd detection in scenes captured by a moving camera.

The camera attached to the aerial vehicle, such as robots, undergoes vibration, and instability, or violent motion. This results in frame blurring and motion measures can become too large or insignificant; the fast motion of camera may result in poor distinction among slower moving objects and people in the scene. These challenges are further complicated due to the distance between the camera and captured the scene as the distance is large, as well as varying over time. The distance between the camera and scene makes the geometry negligible [104, 113]. As the distance between the camera and scene makes the geometry unreliable, it limits the use of methods based on epipolar solutions [41, 104, 107, 113]. However, in some studies, geometry, shape, and camera constraints prove useful to cater for parallax error due to the presence of a perspective angle introduced in the video due to an aerial view of the scene [72, 107, 114, 115].

As homography is used to compensate for camera motion where scene contains parallax. For cases where the camera may rotate or translate, registration of frames is used which removes the effect of zoom, calibration or rotation [116, 117]. For motion compensation, Qian and Medioni [118] used Affine based motion compensation (estimated with RANSAC), while Hu, Chen [47] proposed Ego-motion compensation to obtain more reliable results from the image difference scheme. Qian and Medioni [118] analyzed the scene using geometric constraints (projecting motion in a 4D space), motion flow extraction (using optical flow) and finding flow association. They used Tensor-based voting for analyzing motion's geometric property followed by the flood-fill algorithm for segmentation. Wan Yanli [54], however, suggested that since the camera motion is independent of the object's, motion compensation using affine transform only, for both background and foreground, will not be suitable. So, the pixels used to estimate the affine parameters should only contain the background pixels and they reposed using a two-layer iteration based method as shown in Figure 11.

Figure 11. An example of camera motion compensation is illustrated here [118]

Camera motion compensation based on two-layer iteration was proposed by [54] and is shown in Figure 11. The outer layer iteration is used to update the foreground background feature points using the current affine parameters for detecting the motion regions. Features used are FAST corner-points and edges. The RANSAC process is used to estimate the affine parameters on iteration based on the updated background features.

As fast motion of camera may result in poor distinction among slower moving objects and people in the scene [110]. Perko, Schnabel [66] proposed to use OF gathered from frames with a temporal distance of 10 frames with the purpose of averaging the flow and smoothing the features. Georgiadis, Ayvaci [119] proposed a solution to this problem by using dominant motion patterns and detecting regions that violate the normal motion model by independently moving. These regions are classified as inliers, outliers, and true positions inference to overcome the requirement for long trajectories. DeGol and Nam [7] used the intuition that in a scene, long-term trajectories of moving objects can be discerned reliably from trajectories of long-term stationary objects regardless of camera motion. Crowd Counting and Density Estimation from Moving Airborne Camera: Since the detection based methods would fail from such a distant camera view, it is more logical to analyze holistic level features coupled with regression based crowd-count estimate methods. Counting crowd moving aerial camera view is still an unexplored research area. However, it's applications can be of high variety and significance in different everyday surveillance tasks.

3.2.3 Crowd Detection from Moving Aerial Camera

Aerial videos of crowds from real world scenario datasets. They contain complete information of the entire scene. They are more practical to cater for emergency situations, where assistance and response time can be improved for rehabilitation, monitoring of services and conditions of crowds in situations such as in flood relief and more. In the videos recorded from the UAVs people in the frame appear to be smaller, tilted and may also be occluded from view at times [37]. In addition, the wind causes the UAV to tilt, lose or gain height causing further complexity to activity detection. Hence all these inconsistencies increase the challenge of crowd detection from moving the aerial platform.

An example of how a crowd looks when captured from the aerial camera is shown in Figure 12**.** The crowd is captured in nadir looking images. In all cases, Figure 12 (a),(b) and (c), it is termed as a dense crowd. In reference to crowd dynamics discusses earlier in section 2, it can be seen that an image can visually be described as a dense crowd image based on the distance from the camera, and density per region. In this section, we explore some recent research works that detect crowd from an aerial moving camera.

(a) (b) (c)

Figure 12. (a) An example of a dense crowd standing in front of an open-air stage [34];(b) Another example of a stadium image captured from satellite [120], and (c) a result of crowd detection from satellite image [120].

1) Detection of crowd from above

Figure 12 shows some examples of nadir looking images. In literature, generally, the first step when dealing with moving or static camera, is to detect foreground and background regions.

$$I = \alpha F + (1-\alpha)\beta$$

where α and β represent foreground and background respectively. A recent work of [66] deals with airborne nadir looking images. Their approach extracts local features (FAST) and uses them to estimate the crowd density. The method performs feature selection step to distinguish local features as a part of person or background. The density is extracted using a kernel density estimate based on the occurrence of features. The number of individuals is spatially aggregated for crowd density estimate. The main objective of this approach is to target the crowd density to detect crowded regions in a frame. Another recent research uses UAV based videos of a low-density crowd. Their method first targets the upper body detection using HOG [121] detector. Invariance of HOG detector to the scale of the image influenced its use in this approach, as it makes its applicability more relevant for aerial based applications. This is especially because as aerial vehicle's height is not fixed at an altitude, it causes variation in the scale of the person. The method of [37] then approximated the location and scale of the upper body of all human subjects in the video. Depending on the height of the UAV, the approximate distance of the crowd from the UAV was computed. This distance was used to estimate bounding boxes. Background and crowd based bounding boxes were distinguished using standard deviation of crowd based boxes. Despite an aerial video, this method could also use progressive search space reduction technique, proposed earlier by Ferrari [122] to reduce the search space for human body parts for estimating human pose. A variant of HOG, [123] has also been used in airborne crowd density estimation [124]. To estimate regions belonging to foreground or background, [125] applied a region-growing algorithm to detect large homogeneous regions. This property was associated with regions belonging to the background. However, an interesting approach of [126], tried to reduce the cost of foreground and background block selection by the following equation

$$J(\alpha, a, b) = \sum_{j \epsilon I} \sum_{i \epsilon w_j} (\alpha_i - \alpha_i I_i - b_j)^2 + \varepsilon a_j^2)$$

where w denotes local image window, a and b represent foreground and background, respectively.
Although these measures might still include some clutter, these techniques are nevertheless helpful in further processing and to estimate the regions for the results of crowd detection.

2) Motion Estimation of crowd from above

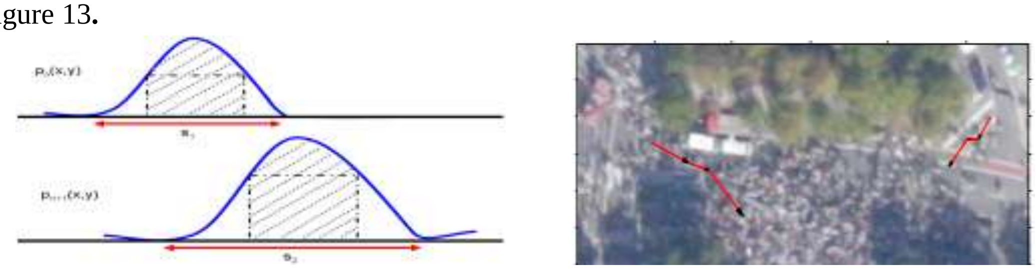
Detecting the motion of people from an airborne platform can be a challenging task. The area covered by each individual can be as small as a few pixels and visual same as a circular dot. The appearance can also change due to the motion of the platform. However, researchers have found some methods to attend to such limitations.

The Optical Flow method, or its variants, can be used to estimate small motions from adjacent video frames from an aerial camera [125]. To ensure a smooth motion flow, frames with a temporal distance of 10 frames have been used in [66]. Because image geometry poses the challenge of how to discern between object motion and camera motion, therefore, [66] transformed the reference 2D image coordinate and the corresponding search 2D images coordinate into 3D world coordinates. These two world coordinates define the real object motion vector independent of the camera movement. Since the temporal difference of the two input video frames is known, the speed of the motion can be calculated in meters per second.

For crowd motion analysis from the airborne video, another work [127] approached this problem by using probability density functions (pdfs) of frames in the sequence. Using Gaussian symmetric kernel functions, they estimated pdfs as

$$p(x, y) = \frac{1}{R} \sum_{i=1}^{K_i} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{(x - x_i)^2 + (y - y_i)^2}{2\sigma}\right)$$

They proposed an automatic kernel bandwidth σ estimation method which was robust to scale and resolution changes, to suit different airborne video platforms. After calculating p(x, y) thresholding was applied to detect regions having high probability values. These regions were labeled as crowded regions. Their work showed that when crowd is moving, the local maximum in pdf will be shifted towards the main movement direction. This idea is illustrated in Figure 13**.**



(a) (b)

Figure 13 Demonstration of crowd motion analysis using pdf functions of images in the sequence, as illustrated in [127]. In 13(a) if the first plot is two-dimensional pdf of nth frame, and bottom one a pdf of n+1[sh] frame in the sequence, they proposed that crowd movement direction can be determined as the direction of a local maximum of the first function, to the closest local maximum of the second function. In (b), estimated main motion directions are plotted on Oktoberfest image in their dataset.

Usually, in the literature of moving camera based detections, the stage of detection of the crowd and feature detection is followed by feature matching. In [126], to detect motion of a human, they used matching pairs technique. To reduce computational complexity, they reduced the number of matching pairs. They used the cue that since airborne videos record negligible human motion, hence finding the same head blob in frame and frame$_{t+1}$ is much probable. Hence, the blob with the highest probability in both frames was the target blob for matching. For a higher density crowd, for instance, medium crowd, comparing each blob pair of size $n \times m$ world be computationally challenging, therefore, they proposed applying temporal analysis on fixed window size, W_t=10, where each window was argued to contain no more than one person. The formula for finding probability of blob c_j being a target for matching frame for motion estimation was calculated as

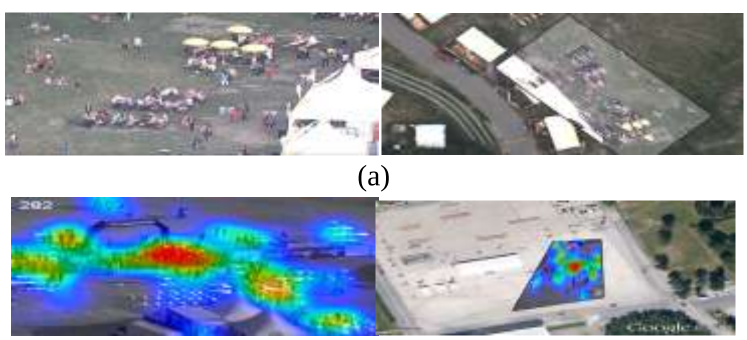
$$P_i(c_j) = P(c_j|r_i)\exp\left(-\frac{D(r_i, c_j)}{\tau}\right)$$

where τ was the constant scaling parameter, and identity with the largest $P_i(c_j)$ is selected for matching.

3) Georeferencing and Homography

Geo-referencing, also called ortho-rectification, is a standard method in photogrammetry and in remote sensing which projects the image onto the earth's surface in a given map projection. Global digital surface models like SRTM1 or ASTER GDEM2 are freely available [66]..The basis for near-real-time mapping is provided with a coupled real-time GPS/IMU navigation system which enables accurate and direct georeferencing. This method generally includes superimposing the video frame with the earlier processed crowd density and with motion map information. This information is geo-referenced and overlayed in Google Earth, as can be seen in Figure 14.

To be able to handle the distortions due to the topography a digital surface model (DSM) is used in [66, 125,128-130]. In their method, the meta-data (GPS/IMU) supplied by the camera system was taken for each frame. Since every frame has a different exterior parameter, it required geo-referencing every frame independently. To georeference the density [66,131,132] projected each density pixel into the common frame. To ensure the human count stays same in the image and world coordinates, they proposed that if any pixel belongs to both of them, the count of these pixels is calculated as the human count [133-135]. Then, 3D world coordinate was calculated.



(a)

(b)

Figure 14 Geo-referencing of a given image for the test site(a) Lakeside, (b) Donauinsel: (bottom) Airborne video frame and (right) the geo-referenced version of (left) overlayed on a true ortho image shown as Google Earth1 overlay. [66]

After locating the features in different frames, the motion model of the camera can be estimated, as discussed in 3.2.2 earlier. In literature, however, because crowd motion is not very high, neither the datasets currently present include a high-speed motion [136-138], therefore this method has not been explored in detail. However, for more information, the reader is redirected to [139-140]]. In the following section, we list different datasets available in the field of crowd detection.

4. BENCHMARK CROWD DENSITY ESTIMATION AND DATASETS

In the literature, some benchmark datasets have been frequently used to train, test, evaluate and validate the crowd detection, estimation, and counting techniques. Table 7 illustrates some of the datasets that have been more repeatedly used:

Table 7: Commonly used Datasets for crowd detection.

Dataset Name	Ref.	Resolution	Color	Place	Crowd	Remarks
FUDAN	[4]	320x240	Grey	Outdoor	3-18	
Grand Central	[19]	720x480	Grey	Indoor	250	
PETS2009	2009	640x480	Grey	Outdoor	3-46	
QMUL	[28]	360x288	RGB	Outdoor		
SAIVT-QUT	[33]	704x576, 352x288	RGB	Indoor	3-23	3 Cam view
Rodriguez	[2]	720x480	RGB	Indoor/ Outdoor		Dense crowd Video.
Mall	[39]	640x480	RGB	Indoor	13-53	
UCF	[43]		Grey	Indoor/ Outdoor	93-2000	Images
USCD	[46]	238x158	Grey	Outdoor	11-46	
UMN	[50]		Gray/RGB	Indoor/ Outdoor	2-15	
AgoraSet	Synthetic Dataset					Simulation of crowd
UT	[52]	360x240, 720x480	RGB	Outdoor	1-10	Action, Aerial, Wide View
BIWI	[56]	640x480	RGB	Outdoor		Mobile Camera
ViSor: 3DPeS	[57]	704x576	RGB	Outdoor	1-200	

From the literature review, we have seen that crowd detection and hitherto research applications have been quite active over the past two decades. There has been an increase of researchers' interest in facilitating the automation of crowd detection using the new computer vision techniques. Automation surprisingly comes with different challenges which still have not been quite resolved. This is especially due to the ever increasing variety of the datasets, such as growing video based dataset. Table 8 enlists the datasets available for aerial imagery and videos. The next section discusses general problems and future possibilities in crowd detection methods.

Table 8: Commonly used aerial Datasets for human/crowd detection.

Dataset Name	Ref.	Remarks
WWW	[20]	Video
UFC	[26]	Video
-	[31]	Image
-	[34]	Image
-	[37]	Video
	[2]	Video
CF UCF	[6]	Video
-	[48]	Video

5. GENERAL PROBLEM AND GAP RESEARCH AREAS IN CROWD DENSITY ESTIMATION TECHNIQUES

In this literature review, we have highlighted the crowd dynamics which are complex and which renders the current research challenged for detection, estimation, and counting of the crowd. A more reliable crowd detection system must be adaptable to the dynamics of a crowd; robust to motion states, spatial distances, the motion state of the camera and varying density of crowd in a scene. Most of the existing works which do not attend to these challenges need to be improved. Table 9 below tabulates the methods and general problems in current work.

Table 9: Crowd detection: challenges and advantages for better computer vision

Motion State	Problems	Application
Crowd		
Static	Spatial Differences, Varying density in a scene, Clutter, Occlusion	Crowd analysis, Counting and Management
Moving	Static Individuals, Slow Speed w.r.t objects such as cars	More general crowd scenario
Dynamic (Static, Moving)	Computationally expensive, illumination changes, motion state variation	Advantages Better detection of crowd's dynamic state of motion. More robust to different motion states.
Camera		
Static	People out of sight, Limited Field of View, Perspective distortion, Viewpoint variance	Images, Videos,
Moving	Varying motion patterns of camera including different challenges, such as jitter, varying speed, altitude variation, illumination changes,	Videos
Dynamic (Static, Moving)	Motion overestimation	Advantages Can get ROI by getting regions of high human pressure. More robust crowd detection scenario.

These methods typically assume that image regions of target objects have a distinct motion and/or appearance, and another moving object cannot be confused with them. Datasets for aerial footage of moving objects mostly consider fast moving objects such as cars. Crowd dynamics vary from other objects and the moving vehicle detection methods cannot be extended to occlusion and perspective distortion prone crowd movements with only a little motion information.

Also, in the current research studies on crowd detection, several challenges in real life applications are missed in benchmark dataset which poses different limitations. Incomprehensive definition of a dense crowd; what is the number of people that constitutes a dense crowd. UCF dataset, for instance, identifies dense crowd images which include thousands of people, whereas Grand Hotel dataset also identifies a dense crowd where the number of people is around 200 only. No established Benchmark for crowd detection from moving platform is established. For either case; static aerial camera or moving aerial camera footage, the research area has many challenges and opportunities which are yet to be explored by the research community.

6. Conclusion

While crowd detection is an oft-explored research area, but several challenges persist and current methods fail to provide a robust solution in case of those challenges. In this paper, challenges faced by current work and applications of the suggested future improvements have been discussed. In this paper, we presented a review on crowd detection, estimation, and counting. This review focuses on crowd detection methods in more challenging scenarios; different crowd motion states, and varying camera motion. For crowd motion states, static and non-static, several methods present in literature are discussed. We also explore limitations in those methods, such as those which oversee the difference in the state of motion. Camera motion state affects the results of current crowd detection systems. While pedestrian detection from a camera mounted on moving the vehicle is a much-studied work, this paper discusses the current work and possibilities of future directions for research in this novel area: crowd detection from an aerial view. Covering these research gaps can help in automating the surveillance, security, and relief based activities.

References

1. Shaheen, K. *Hajj stampede death toll raised to 1,453*. 2015; Available from: http://www.theguardian.com/world/2015/oct/09/hajj-stampede-death-toll-hits-1453.
2. M. Rodriguez, J.S., I. Laptev, and J. Y. Audibert, , *Density-aware person detection and tracking in crowds*, in *IEEE International Conference on Computer Vision*. 2011.
3. Hajer Fradi, J.-L.D., *Towards crowd density-aware video surveillance applications.* Information Fusion, 2015. **24**(3).

4. Ben Tan, J.Z., Liang Wang, , *Semi-Supervised Elastic Net for Pedestrian Counting.* Pattern Recognition,, 2011. **vol. 44,** (issues. 10-11,): p. pp. 2297-2304.
5. G. J. Brostow, R.C. *Unsupervised Bayesian detection of independent motion in crowds*. in *Conference of Computer Vision and Pattern Recognition*. 2006. Washington, DC: ICC.
6. Shah, S.A.a.M., *A lagrangian particle dynamics approach for crowd flow segmentation and stability analysi*, in *IEEE Conference on Computer Vision and Pattern Recognition*. 2007. p. 1-6.
7. DeGol, J. and M. Nam. *A clustering approach for detecting moving objects captured by a moving aerial camera*. in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. 2014. IEEE.
8. B. Zhou, X.W., and X. Tang, *Random field topic model for semantic region analysis in crowded scenes from tracklets,*, in *IEEE Conference on Computer Vision and Pattern Recognition*. 2011. p. 3441–3448.
9. Hashemzadeh, M. and N. Farajzadeh, *Combining keypoint-based and segment-based features for counting people in crowded scenes.* Information Sciences, 2016.
10. B. Zhou, X.W., and X. Tang, *Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents*, in *IEEE Conference on Computer Vision and Pattern Recognition*. 2012. p. 2871–2878.
11. C. Wang, X.Z., Z. Wu, and Y. Liu, *Motion pattern analysis in crowded scenes based on hybrid generative-discriminative feature maps*, in *IEEE International Conference on Image Processing*. 2013. p. 2837–2841.
12. C. Wang, X.Z., Y. Zou, and Y. Liu, *Analyzing motion patterns in crowded scenes via automatic tracklets clustering.* China Communications, 2013. **10**(4).
13. BBC. *Egypt suspends football league after Cairo stadium deaths*. 2012; Available from: http://www.bbc.com/news/world-middle-east-31252429.
14. S. Bak, D.-P.C., J. Badie, E. Corvee, F. Br´emond, and M. Thonnat, *Multi-target tracking by discriminative analysis on riemannian manifold.* IEEE International Conference on Image Processing, 2012: p. 1605–1608.
15. M. Lewandowski, D.S., D. Makris, S. Velastin, and J. Orwell,, *Tracklet reidentification in crowded scenes using bag of spatiotemporal histograms of oriented gradients.* Pattern Recognition, 2013. **7914**: p. 94–103.
16. C.-H. Kuo, C.H., and R. Nevatia, *Multi-target tracking by online learned discriminative appearance models*, in *IEEE Conference on Computer Vision and Pattern Recognition*. 2010. p. 685–692.
17. M. Hu, S.A., and M. Shah, *Detecting global motion patterns in complex videos*, in *International Conference on Pattern Recognition*. 2008.
18. Reisman, P., et al. *Crowd detection in video sequences*. in *Intelligent Vehicles Symposium, 2004 IEEE*. 2004.
19. B. Zhou, X.W.a.X.T., *Understanding Collective Crowd Behaviors: Learning a Mixture Model of Dynamic Pedestrian-Agents.* Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2012
20. Jing, S., et al. *Deeply learned attributes for crowded scene understanding*. in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*. 2015.
21. P. Felzenszwalb, R.G., D. McAllester, and D. Ramanan, *Object detection with discriminatively trained part based models.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010. **32**(9): p. 1627-1645.
22. Donatello Conte, P.F., Gennaro Percannella, Francesco Tufano, Mario Vento *A method for counting moving people in video surveillance videos.* EURASIP Journal of Advance Signal Processing, 2010: p. 231-240.
23. Jalalian, A. and M. Fathy. *Pedestrian Detection from a Moving Camera with an Advanced Camera-Motion Estimator*. in *Signal-Image Technologies and Internet-Based System, 2007. SITIS '07. Third International IEEE Conference on*. 2007.
24. Sims, A. *Hindu festival deaths: Dozens of pilgrims crushed in stampede to retrieve shoes*. 2013; Available from: http://www.independent.co.uk/news/world/asia/hindu-festival-deaths-dozens-of-pilgrims-crushed-in-stampede-to-retrieve-shoes-10386975.html.
25. Ess, A., B. Leibe, and L. Van Gool. *Depth and Appearance for Mobile Scene Analysis*. in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. 2007.
26. UCF, *UCF Aerial Action Data Set*, UCF, Editor. 2009.
27. R. Liang, Y.Z., H. Wang, *Counting crowd flow based on feature points.* Neurocomputing 2014. **133**: p. 377-384.
28. C. C. Loy, T.M.H., T. Xiang, and S. Gong, *Stream-based Joint Exploration-Exploitation Active Learning.* Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012: p. pp. 1560-1567.
29. Su, H., H. Yang, and S. Zheng, *The Large-Scale Crowd Density Estimation Based on Effective Region Feature Extraction Method*, in *Computer Vision – ACCV 2010*, R. Kimmel, R. Klette, and A. Sugimoto, Editors. 2011, Springer Berlin Heidelberg. p. 302-313.

30. Lin, C.-C., *Detecting and tracking moving objects from a moving platform.* 2012.
31. Meynberg, O. and G. Kuschk, *Airborne crowd density estimation.* ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2013. **2**: p. 49-54.
32. R. Mehran, B.E.M., and M. Shah, *A streakline representation of flow in crowded scenes*, in *European Conference on Computer Vision*. 2010. p. 439–452.
33. Ryan, D., Denman, Simon, Sridharan, Sridha, & Fookes, Clinton B. , *Scene invariant crowd counting and crowd occupancy analysis.* Video Analytics for Business Intelligence [Studies in Computational Intelligence, 2012. **409**: p. 161-198.
34. Meynberg, O., S. Cui, and P. Reinartz, *Detection of high-density crowds in aerial images using texture classification.* Remote Sensing, 2016. **8**(6): p. 470.
35. Wong, S.W.a.H.S., *Crowd motion partitioning in a scattered motion field.* IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2012: p. 1443–1454.
36. Ghosh, A.G.a.S., *Object Detection From Videos Captured by Moving Camera by Fuzzy Edge Incorporated Markov Random Field and Local Histogram Matching* IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, 2012. **22**(8).
37. Penmetsa, S.F.M.A.S.S.N.O. *Autonomous UAV for suspicious action detection using pictorial human pose estimation and classification*. 2014. **Vol. 13**, p. 18-32.
38. Liu, W.H.a.Z., *Motion pattern analysis in crowded scenes by using density based clustering,*, in *IEEE International Conference on Fuzzy Systems and Knowledge Discovery*. 2012. p. 1855–1858.
39. K. Chen, S.G., T. Xiang, and C. C. Loy, *Cumulative Attribute Space for Age and Crowd Density Estimation*, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. 2013.
40. B. Zhan, D.N.M., P. Remagnino, S.A. Velastin, L. Xu, *Crowd analysis: a survey.* Machine Vision and Applications, 2008. **19**: p. 345-357.
41. Hu, W.-C., et al., *Moving object detection and tracking from video captured by moving camera.* Journal of Visual Communication and Image Representation, 2015. **30**: p. 164-180.
42. Radke, A.M.C.a.R.J., *Detecting dominant motions in dense crowds.* IEEE Journal of Selected Topics in Signal Processing, 2008. **2**(4): p. 568–581.
43. Haroon Idrees, I.S., Seibert Cody, Mubarak Shah *Multi-source Multi-scale Counting in Extremely Dense Crowd Images*. in *Conference on Computer Vision and Pattern Recognition*. 2013.
44. J. Silveira, J.J., S. Musse, and C. Jung, *Crowd analysis using computer vision techniques.* Signal Processing Magazine, 2010. **27**(5): p. 66-77.
45. Y. Yang, J.L., and M. Shah, *Video scene understanding using multiscale analysis*, in *IEEE International Conference on Computer Vision*. 2009. p. 1669–1676.
46. Vijay Mahadevan, W.L., Viral Bhalodia and Nuno Vasconcelos. *Anomaly Detection in Crowded Scenes*. in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR),*. 2010. San Francisco, CA, .
47. Hu, W.-C., et al., *Effective Moving Object Detection from Videos Captured by a Moving Camera*, in *Intelligent Data analysis and its Applications, Volume I*, J.-S. Pan, et al., Editors. 2014, Springer International Publishing. p. 343-353.
48. ISPRS, *Ikonos data set*, ISPRS, Editor.
49. Saleh, S.A.M., S.A. Suandi, and H. Ibrahim, *Recent survey on crowd density estimation and counting for visual surveillance.* Engineering Applications of Artificial Intelligence, 2015. **41**(0): p. 103-114.
50. UMN, *Monitoring Human Activity* University of Minnesota, Department of Computer Science and Engineering.
51. I. Saleemi, L.H., and M. Shah, *Scene understanding by statistical modeling of motion patterns.* IEEE Conference on Computer Vision and Pattern Recognition, 2010: p. 2069–2076.
52. Ryoo, M.S.a.A., J. K., *Spatio-Temporal Relationship Match: Video Structure Comparison for Recognition of Complex Human Activities",*, in *IEEE International Conference on Computer Vision (ICCV)*. 2009: Kyoto, Japan.
53. L. Song, F.J., Z. Shi, and A. K. Katsaggelos, *Understanding dynamic scenes by hierarchical motion pattern mining,*, in *IEEE International Conference on Multimedia and Expo*. 2011. p. 1-6.
54. Wan Yanli, W., Xifu, Hu Hongpu, *Automatic Moving Object Segmentation for Freely Moving Cameras.* Mathematical Problems in Engineering, 2014. **2014**: p. 11.
55. B. Zhou, X.T., and X. Wang, *Coherent filtering: Detecting coherent motions from crowd clutters,.* European Conference on Computer Vision, 2012: p. 857–871.

56. Pellegrini, S., et al. *You'll never walk alone: Modeling social behavior for multi-target tracking*. in *Computer Vision, 2009 IEEE 12th International Conference on*. 2009. IEEE.
57. D. Baltieri, R.V., R. Cucchiara,, *3DPes: 3D People Dataset for Surveillance and Forensics*. 2011: Arizona, USA.
58. P.-M. Jodoin, Y.B., and Y. Wang, *Meta-tracking for video scene understanding*, in *EEE International Conference on Advanced Video and Signal Based Surveillance*. 2013. p. 1–6.
59. G. K. Still, *Crowd dynamics*. 2000, University of Warwick.
60. Reicher, S., *The psychology of crowd dynamics.* Blackwell handbook of social psychology: Group processes, 2001: p. 182-208.
61. W. Fu, J.W., Z. Li, H. Lu, and S. Ma, *Learning semantic motion patterns for dynamic scenes by improved sparse topical coding.* IEEE International Conference on Multimedia and Expo, 2012: p. 296–301.
62. Helbing, D., et al., *Self-Organized Pedestrian Crowd Dynamics: Experiments, Simulations, and Design Solutions.* Transportation Science, 2005. **39**(1): p. 1-24.
63. Lin, C.-C., *"Detecting and Tracking Moving Objects From a Moving Platform"*. 2012, Georgia Institute of Technology. p. 103.
64. Helbing, D., et al., *Self-organizing pedestrian movement.* Environment and planning B: planning and design, 2001. **28**(3): p. 361-383.
65. Helbing, D., I. Farkas, and T. Vicsek, *Simulating dynamical features of escape panic.* Nature, 2000. **407**(6803): p. 487-490.
66. Perko, R., et al., *Counting people from above: Airborne video based crowd analysis.* arXiv preprint arXiv:1304.6213, 2013.
67. Davies, A.C., J.H. Yin, and S.A. Velastin, *Crowd monitoring using image processing.* Electronics & Communication Engineering Journal, 1995. **7**(1): p. 37-47.
68. Marana, A., et al. *Estimation of crowd density using image processing*. in *Image Processing for Security Applications (Digest No.: 1997/074), IEE Colloquium on*. 1997. IET.
69. Jong Taek, L., C. Chia-Chih, and J.K. Aggarwal. *Recognizing human-vehicle interactions from aerial video without training*. in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*. 2011.
70. N. Amir Sjarif, S.S., S. Mohd Hashim, and S. Yuhaniz,, *Crowd analysis and its applications.* Software Engineering and Computer Systems, 2011. **179**: p. 687-697.
71. Shakeri, M. and H. Zhang. *Detection of small moving objects using a moving camera*. in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. 2014. IEEE.
72. Jinman, K., et al. *Detection and tracking of moving objects from a moving platform in presence of strong parallax*. in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. 2005.
73. Casas, J.R., A.P. Sitjes, and P.P. Folch, *Mutual feedback scheme for face detection and tracking aimed at density estimation in demonstrations.* Vision, Image and Signal Processing, IEE Proceedings -, 2005. **152**(3): p. 334-346.
74. T. Brox, A.B., N. Papenberg, and J. Weickert, *High accuracy optical flow estimation based on a theory for warping*, in *European Conference on Computer Vision*. 2004. p. 25–36.
75. R. Mehran, A.O., and M. Shah, *Abnormal crowd behavior detection using social force model,*, in *IEEE Conference on Computer Vision and Pattern Recognition*. 2009. p. 935–942.
76. X. Wang, X.Y., X. He, Q. Teng, and M. Gao, *A high accuracy flow segmentation method in crowded scenes based on streakline.* International Journal for Light and Electron Optics, 2014. **125**: p. 924–929.
77. H. Su, H.Y., S. Zheng, Y. Fan, and S. Wei, *The large-scale crowd behavior perception based on spatio-temporal viscous fluid field.* IEEE Transactions on Information Forensics and Security, 2013. **8**(10): p. 1575–1589.
78. Y. Benabbas, N.I., and C. Djeraba, , *“Motion pattern extraction and event detection for automatic visual surveillance.* Journal on Image and Video Processing, 2011. **7**: p. 1-15.
79. P.-M. Jodoin, Y.B., and Y. Wang, *Meta-tracking for video scene understanding,*, in *IEEE International Conference on Advanced Video and Signal Based Surveillance,*. 2013.
80. Tomasi, C. and T. Kanade, *Detection and tracking of point features*. 1991: School of Computer Science, Carnegie Mellon Univ. Pittsburgh.
81. Celik, H., A. Hanjalic, and E.A. Hendriks. *Towards a Robust Solution to People Counting*. in *Image Processing, 2006 IEEE International Conference on*. 2006.

82. Kilambi, P., et al., *Estimating pedestrian counts in groups.* Computer Vision and Image Understanding, 2008. **110**(1): p. 43-59.
83. Hashemzadeh, M. and N. Farajzadeh, *Combining keypoint-based and segment-based features for counting people in crowded scenes.* Information Sciences.
84. Ryan, D., et al. *Crowd counting using multiple local features*. in *Digital Image Computing: Techniques and Applications, 2009. DICTA'09.* 2009. IEEE.
85. T. Zhao, R.N., and B. Wu, *Segmentation and tracking of multiple humans in crowded environment.* IEEE Transactions on Pattern Analysis and Machine Intelligence,, 2008. **30**(7): p. 1198–1211.
86. G. J. Brostow, R.C., *Unsupervised bayesian detection of independent motion in crowds*, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2006. p. 594–601.
87. S. Wu, B.E.M., and M. Shah, *Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes,* in *IEEE Conference on Computer Vision and Pattern Recognition*. 2010. p. 2054–2060.
88. P. Allain, N.C., and T. Corpetti, *Crowd flow characterization with optimal control theory*, in *Asian Conference of Computer Vision*. 2009. p. 279–290.
89. Peri´c, J.H.F.a.M., *Computational methods for fluid dynamics.*, ed. Springer Berlin. Vol. 3. 1996.
90. Teng, L., et al., *Crowded Scene Analysis: A Survey.* Circuits and Systems for Video Technology, IEEE Transactions on, 2013. **25**(3): p. 367-386.
91. Kong, D., D. Gray, and H. Tao. *A viewpoint invariant approach for crowd counting*. in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. 2006. IEEE.
92. Siu-Yeung Cho, T.W.S.C., Chi-Tat Leung, *A neural-based crowd estimation by hybrid global learning algorithm.* IEEE Trans. Syst. Man Cybern. Part B:Cybern, 1999. **29**(4): p. 535-541.
93. A. N. Marana, L.D.F.C., R. A. Lotufo, and S. A. Velastin, . *Estimating crowd density with Minkowski fractal dimension,*. in *Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing* 1999.
94. Rahmalan, H., M.S. Nixon, and J.N. Carter. *On Crowd Density Estimation for Surveillance*. in *Crime and Security, 2006. The Institution of Engineering and Technology Conference on*. 2006.
95. Albiol A. , S.J., *Statistical video analysis for crowds counting*, in *16th IEEE International Conference on Image Processing (ICIP)*. 2009. p. 2569-2572.
96. C. Harris. *A combined corner and edge detector*. in *Proceedings of Alvey Vision Conference*. 1988.
97. Shah, S.A.a.M., *Floor fields for tracking in high density crowd scenes.* European Conference on Computer Vision, 2008: p. 1–14.
98. Marco Manfredi, R.V., Simone Calderara, Rita Cucchiara, *Detection of static groups and crowds gathered in open spaces by texture classification.* Pattern Recognition Letters, 2014.
99. Cucchiara, R., A. Prati, and R. Vezzani, *Real-time motion segmentation from moving cameras.* Real-Time Imaging, 2004. **10**(3): p. 127-143.
100. Duong, N., C. Hughes, and J. Horgan. *Optical Flow-Based Moving-Static Separation in Driving Assistance Systems*. in *Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on*. 2015.
101. Jodoin, P.-M., M. Mignotte, and C. Rosenberger, *Segmentation framework based on label field fusion.* Image Processing, IEEE Transactions on, 2007. **16**(10): p. 2535-2550.
102. Wang, Y., *Joint random field model for all-weather moving vehicle detection.* Image Processing, IEEE Transactions on, 2010. **19**(9): p. 2491-2501.
103. Ghosh, A., B.N. Subudhi, and S. Ghosh, *Object detection from videos captured by moving camera by fuzzy edge incorporated markov random field and local histogram matching.* Circuits and Systems for Video Technology, IEEE Transactions on, 2012. **22**(8): p. 1127-1135.
104. Franke, U., et al., *Autonomous driving goes downtown.* Intelligent Systems and their Applications, IEEE, 1998. **13**(6): p. 40-48.
105. Liang, Z. and C. Thorpe. *Stereo- and neural network-based pedestrian detection*. in *Intelligent Transportation Systems, 1999. Proceedings. 1999 IEEE/IEEJ/JSAI International Conference on*. 1999.
106. Broggi, A., et al. *Shape-based pedestrian detection*. in *Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE*. 2000.

107. Wedel, A., et al., *Detection and Segmentation of Independently Moving Objects from Dense Scene Flow*, in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, D. Cremers, et al., Editors. 2009, Springer Berlin Heidelberg. p. 14-27.
108. Brox, T. and J. Malik, *Object segmentation by long term analysis of point trajectories*, in *Computer Vision–ECCV 2010*. 2010, Springer. p. 282-295.
109. Viola, P., M.J. Jones, and D. Snow. *Detecting pedestrians using patterns of motion and appearance*. in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. 2003.
110. Sheikh, Y., O. Javed, and T. Kanade. *Background Subtraction for Freely Moving Cameras*. in *Computer Vision, 2009 IEEE 12th International Conference on*. 2009.
111. Dong, Z. and L. Ping. *Motion Detection for Rapidly Moving Cameras in Fully 3D Scenes*. in *Image and Video Technology (PSIVT), 2010 Fourth Pacific-Rim Symposium on*. 2010.
112. Rita Cucchiara, A.P., Roberto Vezzani, *Real-time motion segmentation from moving cameras.* Real-Time Imaging, 2004. **10**: p. 127–143.
113. Cutler, R. and L.S. Davis, *Robust real-time periodic motion detection, analysis, and applications.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2000. **22**(8): p. 781-796.
114. Pong, W.B.T.a.T.-C., *Detecting Moving Objects.* International Journal of Computer Vision, 1990. **4**: p. 39-57.
115. Nelson, R.C. *Qualitative detection of motion by a moving observer*. in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*. 1991.
116. Seyed Ali Cheraghi, U.U.S., *Moving Object Detection Using Image Registration for a Moving Camera Platform*, in *IEEE International Conference on Control System, Computing and Engineering*. 2012: Penang, Malaysia.
117. Arandjelović, O., D.-S. Pham, and S. Venkatesh, *Efficient and accurate set-based registration of time-separated aerial images.* Pattern Recognition, 2015. **48**(11): p. 3466-3476.
118. Qian, Y. and G. Medioni. *Motion pattern interpretation and detection for tracking moving vehicles in airborne video*. in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. 2009.
119. Georgiadis, G., A. Ayvaci, and S. Soatto. *Actionable saliency detection: Independent motion detection without independent motion estimation*. in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. 2012.
120. Sirmacek, B. and P. Reinartz, *Feature analysis for detecting people from remotely sensed images.* Journal of Applied Remote Sensing, 2013. **7**(1): p. 073594-073594.
121. Dalal, N. and B. Triggs. *Histograms of oriented gradients for human detection*. in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. 2005. IEEE.
122. Ferrari, V., M. Marin-Jimenez, and A. Zisserman. *Progressive search space reduction for human pose estimation*. in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. 2008. IEEE.
123. Felzenszwalb, P.F., et al., *Object detection with discriminatively trained part-based models.* IEEE transactions on pattern analysis and machine intelligence, 2010. **32**(9): p. 1627-1645.
124. Perko, R., et al. *Towards View Invariant Person Counting and Crowd Density Estimation for Remote Vision-Based Services*. in *23rd International Electrotechnical and Computer Science Conference, Portorož, Slovenia. Submitted for review*. 2014.
125. Burkert, F., et al., *People tracking and trajectory interpretation in aerial image sequences.* Photogrammetric Computer Vision and Image Analysis. IAPRS, 2010. **38**: p. 209-214.
126. Yeh, M.-C., H.-K. Chiu, and J.-S. Wang, *Fast medium-scale multiperson identification in aerial videos.* Multimedia Tools and Applications, 2016. **75**(23): p. 16117-16133.
127. Sirmacek, B. and P. Reinartz. *Automatic crowd density and motion analysis in airborne image sequences based on a probabilistic framework*. in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. 2011. IEEE.
128. Dong, J., et al., *Instantaneous video stabilization for unmanned aerial vehicles.* Journal of Electronic Imaging, 2014. **23**(1): p. 013002-013002.
129 Muhsin; Z.F. Rehman, A.; Altameem, A.; Saba, A.; Uddin, M. (2014). Improved quadtree image segmentation approach to region information. the imaging science journal, vol. 62(1), pp. 56-62, doi. http://dx.doi.org/10.1179/1743131X13Y.0000000063.
130 Neamah, K. Mohamad, D. Saba, T. Rehman, A. (2014). Discriminative features mining for offline handwritten signature verification, 3D Research vol. 5(2), doi. 10.1007/s13319-013-0002-3

131 Joudaki, S. Mohamad, D. Saba, T. Rehman, A. Al-Rodhaan, M. Al-Dhelaan, A. (2014) Vision-based sign language classification: a directional Review, IETE Technical Review, Vol.31 (5), 383-391, doi. 10.1080/02564602.2014.961576

132 Soleimanizadeh, S., Mohamad, D., Saba, T., Rehman, A. (2015) Recognition of partially occluded objects based on the three different color spaces (RGB, YCbCr, HSV) 3D Research, Vol. 6 (3), 1-10, doi. 10.1007/s13319-015-0052-9.

133 Younus, Z.S. Mohamad, D. Saba, T. Alkawaz,M.H. Rehman, A. Al-Rodhaan,M. Al-Dhelaan, A. (2015) Content-based image retrieval using PSO and k-means clustering algorithm, Arabian Journal of Geosciences, vol. 8(8) , pp. 6211-6224, doi. 10.1007/s12517-014-1584-7

134 Nodehi, A. Sulong, G. Al-Rodhaan, M. Al-Dhelaan, A., Rehman, A. Saba, T. (2014) Intelligent fuzzy approach for fast fractal image compression, EURASIP Journal on Advances in Signal Processing, doi. 10.1186/1687-6180-2014-112.

135 Ahmad, AM., Sulong, G., Rehman, A., Alkawaz,MH., Saba, T. (2014) Data Hiding Based on Improved Exploiting Modification Direction Method and Huffman Coding, Journal of Intelligent Systems, vol. 23 (4), pp. 451-459, doi. 10.1515/jisys-2014-0007

136 Rehman, A., and Saba, T. (2013) An intelligent model for visual scene analysis and compression, International Arab Journal of Information Technology, vol.10(13), pp. 126-136

137 Basori, AH., Alkawaz, MH., Saba, T. Rehman, A. (2016) An overview of interactive wet cloth simulation in virtual reality and serious games, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, doi. 10.1080/21681163.2016.1178600

138 Alkawaz, M.H., Sulong, G., Saba, T. Rehman, A (2016). Detection of copy-move image forgery based on discrete cosine transform, Neural Computing and Application. doi:10.1007/s00521-016-2663-3

139 Kolivand, H. Sunar, M.S. Rehman, A. Almazyad, A.S. and Saba, T. (2017) Outdoor 3D Illumination in Real Time Environments: A Novel Approach Journal of Information Hiding and Multimedia Signal Processing, vol. 8(1), pp. 208-217.

140 Alsayyh, M.A.M.Y, Mohamad, D. Saba, T. Rehman, A. and AlGhamdi, J.S. (2017) A Novel Fused Image Compression Technique Using DFT, DWT, and DCT, Journal of Information Hiding and Multimedia Signal Processing, vol.8(2), pp. 261-271.

141 Saba, T. Rehman, A. Altameem, A. Uddin, M. (2014) Annotated comparisons of proposed preprocessing techniques for script recognition, Neural Computing and Applications Vol. 25(6), pp. 1337-1347 , doi. 10.1007/s00521-014-1618-9

142 Rehman, A., and Saba, T. (2013) An intelligent model for visual scene analysis and compression, International Arab Journal of Information Technology, vol.10(13), pp. 126-136.

143 Meethongjan,K. Dzulkifli,M. Rehman,A. Altameem,A. Saba, T. (2013) An intelligent fused approach for face recognition, Journal of Intelligent Systems vol.22(2), pp. 197-212.

144 Haron, H. Rehman, A., Wulandhari, L.A., Saba, T. (2011) Improved vertex chain code based mapping algorithm for curve length estimation, Journal of Computer Science vol. 7(5), pp. 736-743.

145 Lung, JWJ., Salam, MSH, Rehman, A., Rahim, MSM., Saba, T. (2014) Fuzzy phoneme classification using multi-speaker vocal tract length normalization, IETE Technical Review, vol. 31 (2), pp. 128-136, doi. 10.1080/02564602.2014.892669

146 Selamat, A. Phetchanchai, C. Saba, T. Rehman, A. (2010). Index financial time series based on zigzag-perceptually important points, Journal of Computer Science, vol. 6(12), pp. 1389-1395, doi. 10.3844/jcssp.2010.1389.1395.

147 Ahmad, AM., Sulong, G., Rehman, A., Alkawaz,MH., Saba, T. (2014) Data Hiding Based on Improved Exploiting Modification Direction Method and Huffman Coding, Journal of Intelligent Systems, vol. 23 (4), pp. 451-459, doi. 10.1515/jisys-2014-0007.

148 Harouni, M., Rahim, MSM., Al-Rodhaan,M., Saba, T., Rehman, A., Al-Dhelaan, A. (2014) Online Persian/Arabic script classification without contextual information, The Imaging Science Journal, vol. 62(8), pp. 437-448, doi. 10.1179/1743131X14Y.0000000083.

149 Younus, Z.S. Mohamad, D. Saba, T. Alkawaz, M.H. Rehman, A. Al-Rodhaan, M. Al-Dhelaan, A. (2015) Content-based image retrieval using PSO and k-means clustering algorithm, Arabian Journal of Geosciences, vol. 8(8), pp. 6211-6224, doi. 10.1007/s12517-014-1584-7.

150 Rehman, A. Kurniawan, F. Saba, T. (2011) An automatic approach for line detection and removal without smash-up characters, The Imaging Science Journal, vol. 59(3), pp. 177-182, doi. 10.1179/136821910X12863758415649

151 Saba, T. Rehman, A. Sulong, G. (2010). Non-linear segmentation of touched roman characters based on genetic algorithm, International Journal of Computer Science and Engineering, vol 2(6),pp. 2167-2172.

152 Al-Ameen, Z. Sulong, G. Rehman, A., Al-Dhelaan, A. Saba, T., Al-Rodhaan, M. (2015) An innovative technique for contrast enhancement of computed tomography images using normalized gamma-corrected contrast-limited adaptive histogram equalization, EURASIP Journal on Advances in Signal Processing:32. doi:10.1186/s13634-015-0214-1.

153 Alqahtani, F.A. Saba, T. (2013) Impact of Social Networks on Customer Relation Management (CRM) in Prospectus of Business Environment, Journal of American Sciences, vol. 9(7), pp.480-486.

154 Shafry, MSM., Rehman, A. Kumoi, R. Abdullah, N. Saba, T. (2012) FiLeDI framework for measuring fish length from digital images, International Journal of Physical Sciences, vol.7(4), pp.607-618.

155 Sulong, G., Rehman, A., Saba, T. (2010) Improved offline connected script recognition based on hybrid strategy, International Journal of Engineering Science and Technology, vol.2(6),pp. 1603-1611.

156 Arafat, S. and Saba, T. (2016) Social Media Marketing Vs Social Media Commerce: A Study on Social Media's Effectiveness on Increasing Businesses' Sales, Journal of Business and Technovation, vol. 4(3), 2016, pp. 134-147.

157 M.S. Amin and A. Noori. Mechanism For Farm Mechanization and Careful Planning Using Geographic Information System (GIS), Journal of Business and Technovation, vol.4(1), 2016, pp. 23-28

158 Patwa, N. Kavuri, S.P. Grandhi, B. Hongrak, K. (2016) Study on Effectiveness of Location-Based Advertising on Food Service Industry in Sydney, Journal of Business and Technovation, vol. 4(3), pp. 112-124.

159 Aljalidi, N. Alshedokhi, M. and Saba, T. (2016) The Impact of TripAdvisor Ratings for Tourisim Advertisement in Saudi Arabia.Journal of Business and Technovation, vol.4(2), pp. 90-95.

160 Elsayed, H.A.G., Alharbi, A.N. AlNamlah, H. Saba, T. (2015) Role of Agile Methodology in Project Management and Leading Management Tools, Journal of Business and Technovation, vol.3(3),pp. 188-193