Troop Ability for Identifying Compulsive Occasion in Crowded Surroundings

Srinivasan.M #1, Suresh.G *2, Ranjitkumar.S #3

Electronics And Communication Engineering, Nandha Engineering College, Anna University, Erode.

[1] *srinekm@gmail.com*

[3] *ranjithnandha007@gmail.com*

[2] *goodspeed041@gmail.com*

Abstract— **This paper focuses on detecting and localizing anomalous occasions in videos of crowded scenes, i.e., divergences from a dominant pattern. Both motion and appearance information are considered, so as to robustly distinguish different kinds of anomalies, for a wide range of scenarios. A newly introduced concept based on troop theory, histograms of oriented troops (HOS), is applied to capture the dynamics of crowded environments. HOS, together with the well-known histograms of oriented gradients, are combined to build a descriptor that effectively characterizes each scene. These appearance and motion features are only extracted within spatiotemporal volumes of moving pixels to ensure robustness to local noise, increase accuracy in the detection of local, non-dominant anomalies, and achieve a lower computational cost. Experiments on benchmark data sets containing various situations with human crowds, as well as on traffic data, led to results that surpassed the current state of the art (SoA), confirming the method's efficacy and generality. Finally, the experiments show that our approach achieves significantly higher accuracy, especially for pixel-level occasion detection compared to SoA methods, at a low computational cost.**

Index Terms— **Troop, ability, crowd, anomaly, traffic.**

I. INTRODUCTION

The widespread use of surveillance systems in roads, stations, airports or malls has led to huge amount of data that needs to be analyzed situation even commercial reasons. The task of automatically detecting frames with anomalous or interesting occasions from long duration video sequences has concerned the research com-munity in the last decade. Occasion, and especially anomaly detection in crowded scenes is very important, e.g. for security applications, where it is difficult even for trained personnel to reliably monitor scenes with dense crowds or videos of long duration. Numerous methods have been proposed to assist in this direction.

We deal with the challenging problem of detecting abnormal patterns in videos of crowded scenes that emerge as spatiotemporal changes, both in motion and appearance. An appearance-related anomaly would be, e.g. a bicycle passing through a crowd. Moreover, sudden changes in velocity, like an abrupt increase of its magnitude and the dispersion of individuals in the crowd are detected, indicating that something unusual and potentially dangerous may have occurred.

In this work we propose a novel method for anomaly detection and localization that incorporates both motion and appearance information. We introduce a descriptor created from Histograms of Oriented Gradients (HOG) to capture appearance, and the newly introduced Histograms of Oriented Troops (HOS), to capture frame dynamics. Troop ability has been used in the past only in the framework of Particle Troop Optimization (PSO) in [1], where PSO optimizes a fitness function minimizing the interaction force derived from the Social Force Model (SFM). However, in our work, troops are used in a very different way: the core idea is to construct a prey based on optical flow values over a specific time window and deploy a compact troop flying over it to acquire accurate and discriminative information of the underlying motion. The agents' motion is determined by forces acting on the troop (Sec. IV), which, unlike [1], do not correspond to the SFM, but are used to determine the troop motion and location.

The experimental section shows that our algorithm outperforms state of the art (SoA) algorithms in accuracy and at a low computational cost. Our contribution can be summarized as follows:

1) Troops are used in an original way, via Histograms of Oriented Troops (HOS) that are introduced to characterize crowd motion for anomaly detection. They lead to credibly filtered flow in videos of crowds, resulting to very few noisy flow values. Thus, troop ability captures the motion of crowded scenes in an efficient way that can be extended to other types of videos.

2) The method can be efficiently applied even when the motion in the crowded scene is non-uniform in space and time, and "anomalies" appear locally in a changing context. This is shown in the experiments of Sec. VI on the complete UCSD dataset, where our method's accuracy for pixel level anomaly detection surpasses the SoA.

II. STATE OF THE ART

Main categories: those that use only motion information to detect an abnormality in the scene, and those that use both appearance and motion information to describe the scene dynamics.

In the first category, Wu et al. [8] use chaotic dynamics in particles' representative trajectories as a means to build a model capable of locating an outlier that moves with a different pattern. Even though this method works for very dense videos where a global motion pattern exists, it is unable to detect local abnormalities that take place in a small region in the frame, or in the absence of a global pattern. Activity recognition based exclusively on trajectories is also proposed. However, this method is only based on motion information, completely ignoring the existence of "interesting" activities that exhibit a typical motion pattern.

A common problem that is encountered by all the methods mentioned earlier is their inability to successfully detect anomalies that move similarly to the "normal" motion pattern, as they rely solely on motion characteristics. A second category of methods tackles this issue by incorporating appearance information as well. One work that stands out in this category is that of [21], that uses mixtures of dynamic textures to describe each 3D cuboid extracted from video sequence and detect temporal and spatial abnormalities. However, the computational cost of that algorithm, around 25 sec per frame, makes it prohibitive for many applications. An improved version of this method, with a lower computational cost, that is similar to ours, is found in [22]: that method's accuracy is also improved, but it still remains lower than ours as the experiments in Sec. VI show.

III. PROBLEM FORMULATION

In this work, we address the problem of detecting dynamically changing anomalies in both space and time in videos with crowds of varying densities. In order to effectively capture these anomalies for a wide range of situations, we incorporate both motion and appearance features. Our algorithm uses data derived from automatically extracted regions of

Even though significant research has taken place on occasion and anomaly detection from static cameras [2], [3] the majority of these works address non-crowded scenes, where detailed visual information can be exploited for each individual. However, real-world surveillance scenarios often involve crowds of people or dense traffic, where such information cannot be easily extracted with traditionally used methods .Existing methods are of

interest (ROIs) instead of entire video frames, so as to only process pixels containing information relevant to the occasion taking place, while at the same time achieving a lower computational cost, fewer false alarms, greater precision and successful spatiotemporal localization of anomalies, both on a global and local scale.

In order to extract the ROIs, we apply background subtraction using weighted moving mean [28], as it has been shown to be robust and reliable, however other SoA background subtraction methods like Gaussian Mixture Models (GMMs) could also be used, leading to equivalent results. We define interest points on a dense grid in the resulting foreground and ROIs are described as rectangular areas of fixed size around each interest point. The size of the ROIs is determined at the beginning of each set of experiments, and depends on the camera viewpoint for each dataset. Due to the static nature of surveillance cameras, the block size needs to be set only once for each camera, or in our case for each dataset, and thus does not affect our algorithm's generality. For the UCSD dataset, a ROI of 20×20 pixels is used, as it is large enough to capture activity/appearance related details, but is not too large, so as to include noisy information in the descriptor.

Once ROIs are extracted, the interest points in them are tracked until the next frames using the KLT tracker, while the foreground grid is continuously updated, with new interest points defined in each new frame's foreground area. The resulting ROIs and the interest points in them are considered informative and are retained if at least 60% of that ROI contains motion, otherwise that interest point and its ROI are considered to be noisy and are ignored. The ROI needs to contain at least 60% moving pixels in order to be as informative as possible; if a ROI contains fewer moving pixels, noisy (motionless) data will also be taken into account, while if it is required to contain more moving pixels, potentially informative interest points may be ignored.

A. Appearance Modeling

In order to extract the appearance characteristics of a video sequence, the Histograms of Oriented Gradients (HOG) proposed in [29] are used, as the HOG descriptor has several advantages over other appearance features: it is color invariant as it uses gray scale images, and is also invariant to illumination and local geometric transformations as a result of the normalization that takes place. At the same time, it effectively captures the local edge and gradient structure, so it can distinguish variations in appearance even in small areas of the image. The implementation of HOG that is adopted is that of [30], as it creates direction invariant HOGs by following a mirroring technique, where mirrored shapes are mapped into the same bin. Direction invariant appearance features (HOGs) decrease intra-class variation, e.g. for walking, which is the predominant activity in human crowds, resulting in similar appearance descriptors for motions in opposite directions. This leads to more robust appearance descriptors that are suitable for the needs of anomaly detection in crowded videos, which can describe, for example, the density or sparseness of a crowd more effectively by ignoring directionality (which is not relevant for appearance).

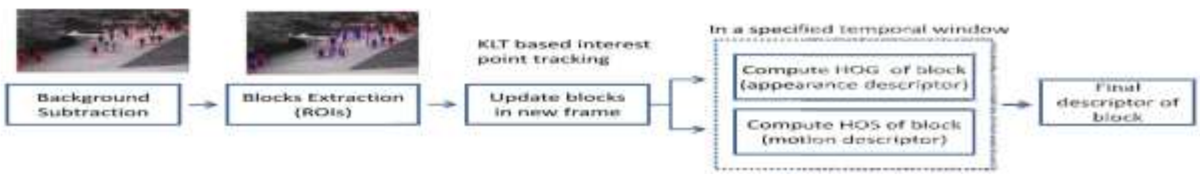
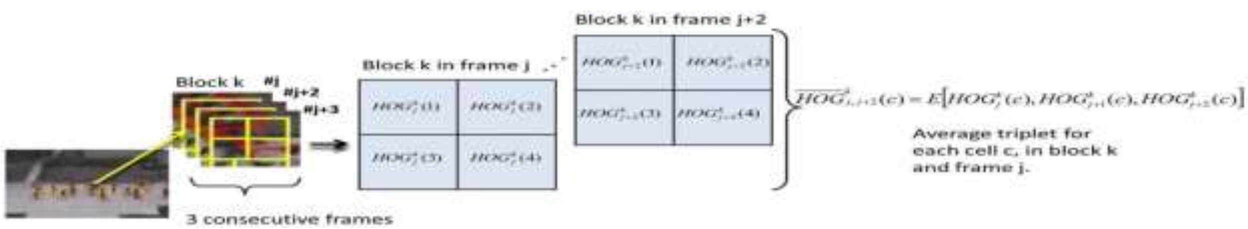


Fig. 1. Problem formulation. (a) Overview of final motion-appearance descriptor calculation.



(b) Extraction of appearance descriptor (HOG).

B. Motion Modeling Using HOS Descriptor

This work introduces a novel method for capturing crowd dynamics based on the application of troop ability, which is used to build a novel motion descriptor. Troop ability in computer science is inspired from the behavior and characteristics of real troops encountered in nature. Troops are comprised of individuals, which act autonomously, while following the specific rules of a troop and interacting with each other. Although the decisions of a troop's individuals take place locally, their aggregated behavior can match occasions in crowded environments, which makes them relevant in many applications.

Troop based methods have been used in the literature for image filtering and noise reduction [31], but their incorporation for the analysis of motion in videos is an original concept first presented in [32]. The core idea is the monitoring of movements in crowded scenes by a troop of agents "flying" over them, to capture their dynamics in a collective way while also taking motion history into account. Troops are thus deployed and the agents' positions are extracted from their accelerated motion, derived from the forces acting on the troop as described in Sec. IV. They are then used to form Histograms of Oriented Troops (HOS), which are used to capture the ROIs' underlying motion and detect anomalous occasions in them. The main concepts of our troop descriptor are presented in the following section.

IV. TROOP MODELLING FOR CROWDS DYNAMICS

In our implementation, we adopt physics-based modeling of crowded scenes, as their properties are highly correlated with those of a troop in nature. The troop model that is used is based on the general theory described in [31] and on the behavior of natural troops, consisting of predators, which "fly" over the "prey", following its dynamics. In [31], troop modeling is used to filter noise in images, whereas in this work it is deployed to better characterize the highly complex and stochastic motion information from videos of crowds. In our implementation, troops comprise of agents and a prey: the agents "track" the prey, but also interact with each other, as they would in nature. Hence, agents ("predators") are subject to three types of forces: "physical" forces, like inertia and friction, interaction forces between them, and external forces dependent on the prey. Interaction forces ensure the cohesion of the troop of agents, friction forces maintain elementary memory of the agents' velocity, while external forces depend on the characteristics of the prey being tracked.

In our case, we are interested in the extraction of motion features via the troop modeling, so optical flow (OF) values are used as a prey, as detailed in the next section. Thus, the use of troops is expected to lead to better results than when using OF information alone, as they can capture the most important aspects of crowd behavior while circumventing the effects of local noise, occlusions and the overall complexity of motion in crowded scenes.

Consequently, in this approach, troop ability maps the motion information into a more informative space by efficiently tracking the motion represented by the prey. Agents filter the prey motion, avoiding false alarms and local noise caused e.g. by occlusions or outlier optical flow values. The prey corresponds to the values of the variable that we want to leverage in the discriminative process. In our case, we are interested in the extraction of motion features via the troop modeling, so optical flow (OF) values are used as a prey, as detailed in the next section. Thus, the use of troops is expected to lead to better results than when using OF information alone, as they can capture the most important aspects of crowd behavior while circumventing the effects of local noise, occlusions and the overall complexity of motion in crowded scenes.

A. Prey Generation

The prey that is tracked by the troop comprises of magnitude values of pixels lying inside ROIs, instead of their luminance, which is the case in [31]. Hence, the number of prey in each frame varies, as it is equal to the number of ROIs in the frame. In this section we describe how prey data is extracted, namely how it is mapped to be tracked by agents. As mentioned previously, ROIs correspond to rectangular areas around each interest point containing a fixed number of *n* pixels. In order to form the prey for a ROI in a temporal window of *m* frames, we consider the pixels of each ROI sequentially over time. Each pixel at position *i* in a particular ROI of frame *j* has OF magnitude equal to O_{ij} , where $1 \le i \le n$ and $1 \le j \le m$.

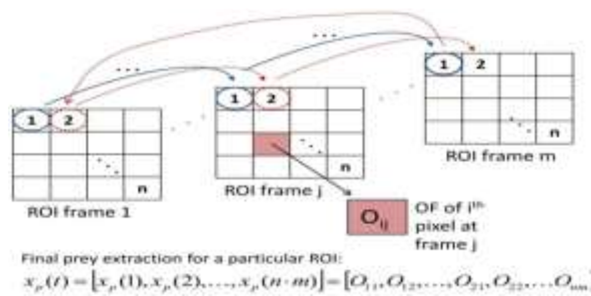


Fig. 2. Prey extraction in a m frame window occurs sequentially in a cuboid of m frames. First, m "OF values" of the 1 st pixel are taken into account, then m instances of the 2 nd pixel and so on, until m instances of the n t h pixel, where n is the number of pixels in each ROI.

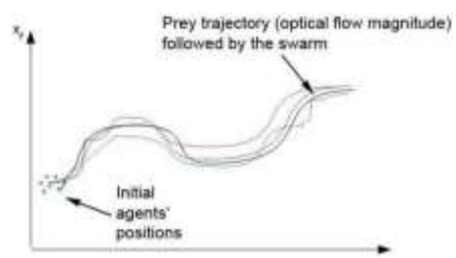
For the prey construction, we consider the i^{th} pixel's OF sequentially over time. The OF magnitude is used to determine the prey's position x_p as follows:

$$x_p(t) = O_{i\,j}$$

Where t is a spatiotemporal index that spans all n ROI pixels over m frames, so that $1 \leq t \leq n \cdot m$. The selection of the sequence of pixels for prey construction is very important for capturing meaningful temporal information.

B. Extraction of Forces

In this section we present the manner in which the agents operate, i.e. the way they "fly over" the prey and track it. Agents are groups that we define to track the prey and characterize its state: they are initially located in random positions, which change over time according to agent-prey forces, agent-to-agent forces and friction forces presented here. The result of these forces' interactions is the accelerated motion of the agents, which is affected and formed according to prey behavior. These forces are inspired by crowd psychology and the analysis of movements of individuals in crowds [33], matching real world behaviors of people (or other entities, like cars or animals) in crowded situations: for example, when agents are too close to each other, repulsive forces develop between them, while the opposite occurs (attraction forces develop) when they are at a large distance, ensuring the cohesion of the troop of agents. An illustrative example of the way the troop follows the prey is given in Fig.3.



C. HOS Descriptor

In order to form the HOS descriptor, we examine the evolution of the agents' positions, determined by prey motion patterns and the forces affecting the agents. We modify Newton's second law of motion by inserting an elementary parameter γ that takes into account the previous velocity values, as in Eq. (9) shown below. Then, the acceleration $\ddot{x}_i(t)$ of each agent i at position $x_i(t)$ is given by the vector sum of all forces acting on it, considering the fact that an agent's mass equals 1, along with the γ-weighted velocity of the previous time instant. Thus, the acceleration of each agent is given by:

$$\ddot{x}_i(t) = (\gamma - 1)\dot{x}_i(t-1) + F_{neigh}(i, t) + F_{fric}(i,t) + F_{ext}(i,p,t),$$

$$\dot{x}_i(t) = \gamma \cdot \dot{x}_i(t-1) + \delta \cdot \ddot{x}_i(t),$$

Where δ constitutes a time step parameter,

During training, ROIs are extracted and the pixel OF in them is examined and tracked by the agents. We then compute the average of troop agents' positions of Eq. (11) for each t, and follow a process similar to the HOG extraction of Sec. III-A to extract weighted histograms of agents' positions (HOS), according to the corresponding OF orientation. As in Sec. III-A, each ROI (block) around each interest point, is partitioned into 2 × 2 cells, and the positions of the troop agents that follow this particular block establish a weighted histogram of 18 bins according to the OF orientation in each cell.

Subsequently, these 4 histograms are concatenated to form the block's HOS. In order to include temporal information, the final motion descriptor contains histograms of subsequent frames, averaged in triplets over each time window.

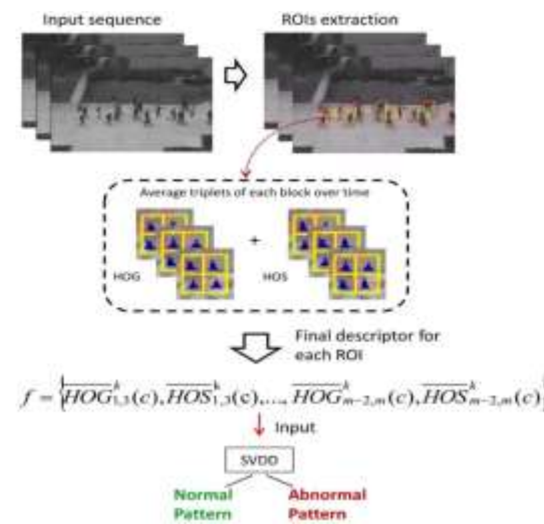


Fig. 3. Overview of method proposed in a time window of m frames.

V. ANOMALY DETECTION AND LOCALIZATION

Appearance and motion descriptors are combined to form the final descriptor for anomaly detection. In a time window of m frames, average triplets of HOG and HOS are consecutively concatenated, resulting in the feature vector of Eq. (12):

$$f = \{HOG_{1,3}, HOS_{1,3}, \ldots, HOG_{m-2,m}, HOS_{m-2,m}\}$$

The Support Vector Data Description (SVDD) method of [34] was chosen, as it is known to be best suited for outlier detection. According to this approach, spherical boundaries are used instead of planar ones around the provided data of the training set. The goal is to enclose nearly all n training examples in a hyper sphere with center o and the smallest possible radius R, with the outliers lying outside this sphere. Thus, its purpose is to minimize the function:

After training, localization is straightforward, as descriptors are estimated spatially in specific ROIs around interest points. Our algorithm checks each frame's ROI independently, infers about its normality and then notifies the system.

Hence, our method is capable of dealing with non-uniformly moving and evolving crowds, as the descriptors are examined and characterized separately in each ROI. It can accurately localize different anomalies in a wide range of videos, from human crowds to traffic, as the experiments that follow demonstrate.

VI. EXPERIMENTS

In order to evaluate the effectiveness of our method, we applied it on four benchmark datasets of surveillance where different kinds of anomalies were detected. Our algorithm's speed and accuracy on a frame and pixel level were calculated and compared with the SoA, demonstrating its effectiveness.

The length of the window should be large enough to contain sufficient information and, at the same time, as small as possible to avoid undesirable delays during the detection process. Hence, we use a temporal window length that depends on the frame rate and the underlying dominant motion, which in our case is the mean walking frequency of a pedestrian. As an example, Fig.5 depicts the optical flow values of a pedestrian for the "ped2" dataset.

A. Sensitivity Analysis

As the extensive analysis proved, after defining the range of each variable that ensures system stability, our algorithm's performance is not particularly influenced by further variations in these parameter values. Finally, the number of agents forming the troop is fixed to 5, as it is empirically found that this number sufficiently represents the filtered motion dynamics of the scene without negatively affecting the algorithm's speed. Experiments showed that the use of more agents heavily increased the computational cost, with a computational time of 3.86 sec per frame if the number of agents increased to 50, while the algorithm's performance actually decreased. This can be attributed to the fact that the presence of too many agents may lead to noisy internal forces due to the density of the troop, which occasionally degrades the results. On the other hand, the use of fewer agents, as few as 2 agents for example in "ped1", also decreased algorithm's performance from 78.87% to 73.66%. The initial agents' speed and accelerations are set to zero, whereas their initial positions are randomly generated.

B. Evaluation Criteria

In order to evaluate our method, we use the same criteria as the SoA literature for benchmark datasets. Thus, the frame and pixel level criteria described in [21] are adopted for UCSD dataset in Sec. VI-C, while the Area under the Curve (AUC) is used for the UMN and U-turn videos described in Sec. VI-D and Sec.VI-E respectively.

The frame level criterion localizes changes only in time, predicting which frames contain an anomaly, without finding its spatial location: a frame is thus characterized as abnormal if it contains at least one abnormality, wherever it is located.

In contrast, the pixel level criterion includes both temporal and spatial anomaly localization, and is used in the literature [21] as follows: if at least 40% of all anomalous pixels are found (as determined by the ground truth annotation), the detection is considered successful and the frame is characterized as abnormal.

True positives and false positives are then derived by comparing the spatiotemporally detected anomalies with the ground truth, leading to Receiver Operating Characteristic (ROC) curves of true positives vs. false positives to evaluate the method's performance. The Equal Error Rate (EER) corresponds to the frame level criterion, while the Detection Rate (DR) corresponds to the pixel level criterion. These metrics have been widely used in the literature for the benchmark UCSD dataset, as they provide a reliable criterion to evaluate method's performance and to compare it with other SoA works. The EER corresponds to the error rate of a system when the false positives (detections of anomalies in a normal situation) are equal to the false negatives (missed anomaly detections).

C. UCSD Dataset

The UCSD dataset is comprised of two subsets "ped1" and "ped2", containing different scenes recorded from different camera angles [21]. Each "ped1", "ped2" subset is divided into a training set containing exclusively normal frames and a test set, including different kinds of anomalies. The dataset consists of crowds of medium density traversing the scene ("ped2") or moving towards and away from camera, adding some perspective ("ped1"). The UCSD dataset constitutes a challenging dataset, as it contains many occlusions, a variety of anomalies, sometimes co-occurring in the same frame, and its resolution is of low quality. Anomalies present in the test set include bicycles, skaters or other wheeled objects moving with different speeds and passing through the crowd, which are in some cases difficult to detect even for human observers.

As can be seen, different kinds of "anomalies" are successfully localized even when they co-occur in the same frame, as in Fig.11(f). A remarkable achievement of the proposed method is that deviations from normal patterns can be also detected in highly occluded scenes.

D. UMN Dataset

The UMN dataset [35] consists of 7739 frames of 320×240 pixels in 3 different scenes (umn_1, umn_2, umn_3) including respectively 2, 6 and 3 scenarios of crowd escape occasions. The first frames in each occasion depict a normal crowd situation, with people walking or standing in the scene, while "anomaly" takes place with a sudden evacuation. This data is quite straightforward, as the "anomaly" is global and can be easily detected even by only using the average frame motion. As a result, many methods have been proposed for this data, achieving near perfect scores.

The main drawback of this dataset is its limited size, in combination with the absence of a separate training set. The limited number of training frames results in a not well defined "normal" class. Our descriptor uses detailed appearance and motion information; however these change significantly, even in the "normal" frames, so it requires more training data for a better defined description of the "normal" occasions. As a result of its limited size, this data does not allow us to demonstrate the true potential of our method, which uses many complex features so as to be applicable to more difficult videos.

Fig. 4. Different kind of anomalies (shown in red blocks)

Fig. **5.Our results for Love Parade. The truck and ambulance are successfully detected going through a very dense crowd. People from the crowd climbing over the railings are also detected as anomalous behaviour.**

E. U-Turn

In order to confirm our method's robustness, we also applied it to a non-crowd dataset. We used the U-turn dataset of [36], which shows normal traffic in a crossroad and some cars making illegal U-turns ("anomaly"). The dataset comprises of 6117 frames of 360 × 240 pixels. The scenes are quite sparse and, in combination with the dataset's limited size, there is not much training data. However, even with limited training samples, all anomalies are perfectly detected and localized, as can be seen in Fig. 15. It is remarkable that in the first frame in Fig. 15, our algorithm correctly distinguishes between an illegal turn and a legal one. Around 3400 frames depicting normal traffic were used for training, and the rest were used for testing. In Fig. 16, the ROC curve of our method for the U-turn data is compared with the results provided by [22]. As it is shown, we achieve the highest AUC at the frame level, equal to 95, 31%, with all "anomalies" having been correctly detected and localized.

F. Love Parade

The algorithm was also tested on the surveillance data of Love Parade 2010 [37], which contains videos of high density crowds. Snapshots are provided in Fig. 17 and, as it can be observed, deviations from normal crowd patterns are correctly detected and localized, despite the little motion present, and the high number of occlusions, due to the high crowd density. Around 1000 frames were used for training with the rest of the frames used for testing. The truck and ambulance are successfully detected while traversing a highly dense crowd, whereas people in the crowd jumping over railings are also detected as an anomalous behavior.

VII. CONCLUSION

In this work, we propose a novel framework for anomaly detection in different scenarios, recorded from static surveillance cameras. Troopability is exploited for the extraction of robust motion characteristics and together, with appearance features form a descriptor capable of effectively describing each scene. Its remarkable performance in 4 completely different kinds of datasets proves the method's generality and its applicability in real life situations. The high detection rate in the UCSD dataset, that greatly out-performs various state-of-the-art approaches, especially on the most challenging pixel level criterion, demonstrates that the proposed algorithm can be effectively used for challenging crowd videos with many occlusions, local noise and local scale variations. This fact in combination with its low computational cost and its effectiveness in different environments, make our algorithm very appropriate for a variety of surveillance applications.

REFERENCES

[1] R. Raghavendra, A. Del Bue, M. Cristani, and V. Murino, "Optimizing interaction force for global anomaly detection in crowded scenes," in *Proc. IEEE Int. Conf.Comput.Vis.Workshops (ICCVW)*, Nov. 2011, 136–143.

[2]. N. P. Cuntoor, B. Yegnanarayana, and R. Chellappa, "Activity modeling using occasion probability sequences," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 594–607, Apr. 2008.

[3]. L. Wang, Y. Qiao, and X. Tang, "Latent hierarchical model of tem-poral structure for complex activity classification," *IEEE Trans. ImageProcess.*, vol. 23, no. 2, pp. 810–822, Feb. 2014.

[4]. W. Hu, X. Zhou, W. Li, W. Luo, X. Zhang, and S. Maybank, "Active contour-based visual tracking by integrating colors, shapes, and motions," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1778–1792, May 2013.

[5]. A. Basharat, A. Gritai, and M. Shah, "Learning object motion patterns for anomaly detection and improved object detection," in *Proc. IEEEConf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.

[6]. C. Piciarelli, C. Micheloni, and G. L. Foresti, "Trajectory-based anom-alousoccasion detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 11, pp. 1544–1554, Nov. 2008.

[7]. D. Tran, J. Yuan, and D. Forsyth, "Video occasion detection: From sub-volume localization to spatiotemporal path search," *IEEE Trans. PatternAnal. Mach. Intell.*, vol. 36, no. 2, pp. 404–416, Feb. 2014.

[8]. S. Wu, B. E. Moore, and M. Shah, "Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes," in *Proc. IEEE Conf.Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, 2054–2060.

[9]. J. C. Nascimento, M. A. T. Figueiredo, and J. S. Marques, "Activity recognition using a mixture of vector fields," *IEEE Trans. ImageProcess.*, vol. 22, no. 5, pp. 1712–1725, May 2013.

[10]. R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *Proc. IEEE Conf. Comput. Vis.Pattern Recognit. (CVPR)*, Jun. 2009, pp. 935–942.

[11]. X. Cui, Q. Liu, M. Gao, and D. N. Metaxas, "Abnormal detection using interaction energy potentials," in *Proc. IEEE Conf. Comput. Vis. PatternRecognit. (CVPR)*, Jun. 2011, pp. 3161–3167.

[12]. H. Ullah and N. Conci, "Crowd motion segmentation and anomaly detection via multi-label optimization," in *Proc. IEEE Int. Conf. PatternRecognit. Workshop (ICPRW)*, Nov. 2012.

[13]. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnor-mal occasion detection," in *Proc. IEEE Conf. Comput. Vis. PatternRecognit. (CVPR)*, Jun. 2011, pp. 3449–3456.

[14]. V. Saligrama and Z. Chen, "Video anomaly detection based on local statistical aggregates," in *Proc. IEEE Conf. Comput. Vis. PatternRecognit. (CVPR)*, Jun. 2012, pp. 2112–2119.

