Anomaly Detection In Crowded Scenes Using Dense Trajectories

Ke Ma, Michael Doescher, Christopher Bodden
Department of Computer Sciences
University of Wisconsin-Madison
{kma, doescher, cbodden}@wisc.edu

Abstract

Abnormal crowd behavior has become a popular research topic in recent years. This is related to a rise in the need for electronic video surveillance. Many methods have been proposed to detect abnormalities, but these methods rely on optical flow or classical classification techniques. We propose to follow the general pipeline used by previous works, but upgrade several components with state-of-the-art techniques. Specifically, we use dense trajectories in place of optical flow, robust features such as social force, HOG, HOF, and MBH, and a single-class support vector machine. We achieve significant improvements in abnormality detection when compared with prior works.

1. Introduction

Interest in crowd analysis continues to grow within the computer vision community. This focus is motivated by the ubiquity of surveillance cameras in crowded areas and their various applications such as crowd monitoring and modelling. One of the typical applications is to detect deviations from normal crowd behaviors, which are referred to as anomalous or abnormal events. Manually monitoring such a large number of surveillance video streams takes considerable labor, and people are most interested in viewing those videos with potentially anomalous events. Therefore, a reliable automatic anomaly detection technique, which can identify video frames with abnormal crowd behaviors from those with normal crowd behaviors, is in urgent need.

Conventional techniques for understanding individual actions involve object detection, segmentation, tracking, etc. However, crowd analysis has been shown to be more challenging because of the high object density, which introduces the problems of severe inter-object occlusion, small object size and similar object appearance that makes conventional techniques infeasible. Another difficulty comes from the complexity of crowd behaviors which involve goal-driven activities, interactions with the environment as well as other people, and various emergent behaviors.

Even the task of anomaly detection itself is not well defined. This is because crowd behaviors are usually context-sensitive. What is viewed as normal in one scenario may become abnormal in another scenario. For example, running around a racetrack is a normal activity, but running on a plaza is often viewed as abnormal; the people may be fleeing from a threat. In general, normal behavior means doing the proper things in the proper context, and that's why crowd behaviors are usually studied according to the scenario of interest. However, it is extremely difficult for computers to learn to understand the scenarios. Furthermore, the available datasets for anomaly detection tend to be biased; there are many more videos with normal behaviors than those with abnormal behaviors. Thus, it is more appropriate to model this task as a novelty detection problem instead of a supervised learning problem.

A great diversity of approaches have been proposed for anomaly detection. Based on the method used to model the motion information, the approaches can be categorized into two classes. The first class is comprised of object tracking and trajectory modeling [26, 21, 7, 4, 27, 33], which can be quite difficult for crowded scenes. The second class adopts alternative representations to avoid object tracking such as optical flow or some other forms of gradient field [15, 3, 16, 22, 32]. Besides motion information, some approaches account for appearance changes to detect anomalies [6, 17, 20, 19, 8, 24, 28].

In this paper, we introduce a computer vision based method for detecting abnormal crowd behavior using dense trajectories [29, 30, 31]. In a framework similar to [22], we place a dense grid of particles over the images and track their motions via dense trajectories, which can be effectively computed from optical flow. With dense trajectories, we then compute features including social interaction forces and employ a bag-of-words approach to build histograms for each video clip. A single-class support vector machine [25] is used to model the clips with normal behaviors and predict whether the test clips contain anomalous events or not.

2. Related Work

Abnormal crowd behavior has become a popular research topic in recent years. There are two main categories of approaches for understanding crowd behaviors. The conventional approaches, usually referred to as object-based approaches, require segmenting and tracking each individual in order to model crowd behaviors [26, 21, 7, 4, 27, 33]. However, this approach is infeasible in densely crowded scenes, where severe occlusions affect the performances of object detection and tracking.

The other approach considers the crowd as a global entity instead of a collection of individuals. Itti and Baldi [15] extract low-level feature descriptors and model normal activity with Poisson statistics. Adam et al. [3] maintain histograms of local optical flow at multiple monitors. Kim and Grauman [16] learn normal patterns of activity locally with a mixture of probabilistic Principal Component Analysis (PCA) models and detect abnormal patterns globally with a space-time Markov Random Field (MRF) model. Mehran et al. [22] leverage the generalized Social Force model from studies of pedestrian behavior modeling [14, 13]. They compute interaction force flow in crowds from optical flow, which is then used as input to build a latent Dirichlet allocation (LDA) model [5] for anomaly detection. Some improvements have been made by introducing particle swarm optimization (PSO) [23], interaction energy potentials [9], and social attribute-awareness [34]. Wu et al. [32] define a chaotic invariant to describe abnormal events. Some works use even more complete representations that take into account both motion and appearance. Boiman and Irani [6] employ spatio-temporal patches and try to reconstruct future patches using data from previous frames. Kratz et al. [17] propose a method to model statistics of spatio-temporal gradients with a coupled HMM. Mahadevan et al. [20, 19] model the normal crowd behavior by mixtures of dynamic textures. Cong et al. [8] detect abnormal events via a sparse reconstruction over the normal bases. Saligrama and Chen [24] detect anomalies based on their local spatio-temporal signatures using statistical aggregates. Wang et al. [28] adopt high frequency information to capture the dynamic properties of spatio-temporal cuboids.

2.1. Dense trajectories

Dense trajectories [29, 30, 31] describe the apparent motion of points in a video sequence using closely packed sample points. They provide good coverage of the entire video including both foreground actions and background movements due to camera motion. Dense trajectories are more robust to irregular abrupt motions and capture complex motion patterns more accurately than the KLT tracker.

The algorithm for computing dense trajectories relies on densely sampling points from each frame and then using information from the optical flow fields to track the points and generate associated trajectories. Specifically, points on a grid with a spacing of 5 pixels in 8 spatial scales are sampled. Each of these points is tracked for 15 frames using the optical flow fields computed by the Farnebäck's algorithm [12]. Each point $P_t = (x_t, y_t)$ at frame t is tracked to the next frame $t+1$ by median filtering in the dense optical flow field $\omega = (u_t, u_t)$:

$$P_{t+1} = (x_{t+1}, y_{t+1}) = (x_t, y_t) + (M * \omega)|_{(\bar{x_t}, \bar{y_t})}, \quad (1)$$

where M is the median filtering kernel and $(\bar{x_t}, \bar{y_t})$ is the rounded position of (x_t, y_t).

Points in homogeneous regions are initially rejected and not tracked; points that do not move are also rejected as their trajectories do not provide information about activity in the scene; and finally those that move too much are also rejected as they are likely to be erroneous.

Trajectory-aligned descriptors such as HOG, HOF and MBH can also be computed once the trajectories are acquired. HOG (histograms of oriented gradients) [10] and HOF (histograms of optical flow) [18] capture static appearance and absolute local motion information respectively, while the MBH (motion boundary histogram) [11] descriptor encodes relative motion and is more robust to camera motion.

Dense trajectories have been shown to be more successful in action recognition than KLT trackers, and are a good candidate for video representation. In this paper, we adopt dense trajectories as the low-level descriptor to capture the motion information in crowded scenes.

2.2. Generalized Social Force model

Inspired by previous socio-psychological studies, Helbing et al. [14] propose the Social Force model to describe the motion of pedestrians. Their later work [13] integrates the herding effects caused by social panic into the model, which is referred to as the generalized Social Force model. The generalized Social Force model suggests that normal crowd behavior is the result of self-organized behaviors of a collection of individuals each with their own goals, and that abnormal crowd behavior such as panic is the result of a shift in the goals of the individuals to escape a particular threat.

Specifically, individuals normally have their own desired destination and attempt to make progress directly toward that goal at their personal desired velocity v_i^p. This is characterized by the personal desired force

$$F_p = \frac{1}{\tau}(v_i^p - v_i), \quad (2)$$

where v_i is the actual velocity of pedestrian i and τ is the relaxation parameter.

At the same time, they are also subject to interaction forces with their local environment. These interaction forces comprise repulsive and attractive forces F_{ped} as a result of the psychological tendency to maintain a socially acceptable distance between strangers and remain close to people they are related or attracted to, as well as environmental forces F_w to avoid hitting walls or other obstacles. Therefore, the interaction force is defined as

$$F_{int} = F_{ped} + F_w. \tag{3}$$

As a consequence of the two types of forces, each pedestrian i with mass of m_i changes his velocity v_i as

$$m_i \frac{dv_i}{dt} = F_p + F_{int}. \tag{4}$$

Panic scenarios in which people attempt to escape from a threatening situation result in the emergence of herding behaviors. Herding behaviors encourage each pedestrian to move with the average velocity of his neighboring pedestrians $\langle v_i^c \rangle$. To incorporate herding behaviors into the above model, the personal desired velocity v_i^p is replaced with

$$v_i^q = (1 - p_i) v_i^p + p_i \langle v_i^c \rangle, \tag{5}$$

where p_i is the panic weight parameter.

Mehran et al. [22] introduce the generalized Social Force model to the computer vision community, and show promising results in anomaly detection. They place a grid of particles over the frames as the representation of high density crowds. In this way, they can track the particles with the underlying optical flow field and avoid tracking individual pedestrian. Under several assumptions, the interaction forces can be estimated and used for detecting abnormal behaviors. In this paper, we propose a method based on this idea and integrate dense trajectories to improve the accuracy and robustness of particle tracking.

2.3. Single-class SVM

The problem of detecting abnormal behavior is best modeled as novelty detection. Given some normal examples we would like to detect future examples which fall outside the normal model. [22] proposes using latent Dirichlet allocation (LDA) [5] to build a topic model for normal examples and then use a thresholding approach based on the log-likelihood that a new example belongs to that topic model. However, this method involves an arbitrary choice of L topics and an arbitrary threshold. We avoid LDA for these reasons.

Schölkopf et al.[25] propose a method for detecting novelties by extending a support vector machine (SVM) to classify a small region containing most of the training data as positive and everywhere else as negative. Formally, for a training set $x_1, \ldots, x_\ell$, the single-class SVM solves the following quadratic program:

$$\min_{w \in F, \xi \in \mathbb{R}^l, \rho \in \mathbb{R}} \frac{1}{2} \|w\|^2 + \frac{1}{\nu \ell} \sum_i \xi_i - \rho \tag{6}$$

$$s.t. \left(w \cdot \Phi\left(x_i\right)\right) \geq \rho - \xi_i \;,\; \xi_i \geq 0.$$

The classification for a new example x is then

$$f\left(x\right) = sign\left(\sum_i \alpha_i k\left(x_i, x\right) - \rho\right), \tag{7}$$

where $k\left(x, y\right)$ is a kernel function and the coefficients α_i are found by solving the dual:

$$\min_{\alpha} \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j k\left(x_i, x_j\right) \tag{8}$$

$$s.t. \; 0 \leq \alpha_i \leq \frac{1}{\nu \ell} \;,\; \sum_i \alpha_i = 1.$$

3. Our Method

In our method, shown in Figure 1, we first decompose the input videos into clips of 16 frames with 1 frame overlap, and resize them to the fixed width of 320 pixels. A proportion of the clips containing normal crowd behaviors are selected as the training set, and all clips are viewed as the testing set.

We then extract dense trajectories using the code provided by [29, 30]. We make several modifications to their code in order to adapt dense trajectories to the anomaly detection task. First, we output sequences of point coordinates rather than the normalized trajectory descriptors to retain the location and magnitude information. Second, we lower the threshold for rejecting stationary trajectories from $\sqrt{3}$ to 0.5 to avoid losing subtle motion information. Third, we only use the trajectories at the original spatial scale so all trajectories are comparable in terms of magnitude. Along with dense trajectories, we also calculate trajectory-aligned

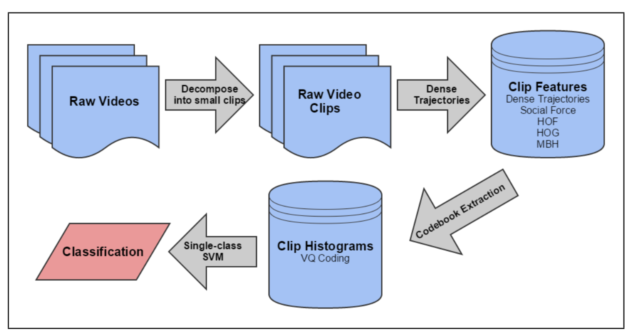

Figure 1. Diagram of the pipeline for our method. Raw videos are decomposed into small clips. Features are extracted from the clips using the dense trajectory method. A codebook is created and a histogram is built for each clip by matching each clip feature to its nearest neighbor in the codebook. Finally the histogram for each clip is used for classification.

descriptors including HOG, HOF and MBH for further investigation.

We make several assumptions to estimate interaction forces from dense trajectories. We view the interaction force F_{int} as a single quantity regardless of its F_{ped} and F_w components. We assume that mass $m_i = 1$ without loss of generality, as the pedestrians we are interested in are of consistently similar sizes for a given scene. For the panic weight parameter p_i, we set it to be 1 in hope that the estimated interaction forces capture herding behaviors in abnormal scenes. Without other prior information, we arbitrarily set the relaxation parameter τ to be 1, which empirically yields good results. Under the above assumptions, the estimation of interaction forces is straightforward. The actual velocity v_i and acceleration $\frac{dv_i}{dt}$ at time t are calculated with particle locations in neighboring frames as

$$v_i = \frac{1}{2}(P_{t+1} - P_{t-1}), \tag{9}$$

$$\frac{dv_i}{dt} = P_{t+1} - 2P_t + P_{t-1}. \tag{10}$$

Then the interaction force F_{int} for this particle at time t is calculated as

$$F_{int} = (\langle v_i^c \rangle - v_i) - \frac{dv_i}{dt}, \tag{11}$$

where $\langle v_i^c \rangle$ is the average velocity of the particle's K nearest neighbors. In our experiments, we choose $K = 5$. Figure 2 shows the visualization of the magnitude of the computed interaction forces overlaid over two video frames.

As each video clip contains 16 frames, we can compute a sequence of 14 interaction forces for each trajectory in the clip. Note that the video clips normally consist of different numbers of trajectories. Therefore, we adopt a bag-of-words approach. We run K-means clustering on all of the interaction force sequences in the training clips to build a codebook of vocabulary size $C = 1024$. Here we only use the magnitude of interaction forces but discard the orientation information, which gives us better results in practice.

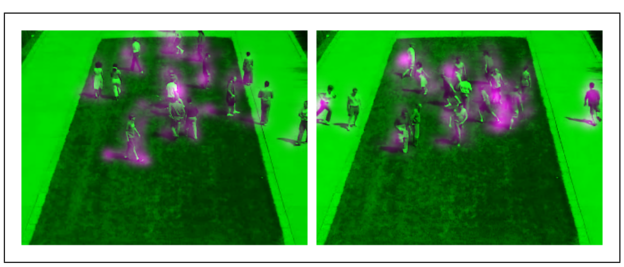

Figure 2. Visualization of the magnitude of the computed interaction forces overlaid over two video frames. The underlying frames in the left and right image are from video clips containing normal and abnormal crowd behaviors respectively.

All interaction force sequences are mapped to codeword indices using vector quantization (VQ) coding according to the codebook. Other coding methods such as locality-constraint linear coding (LLC) were tested but failed to improve the performance. A histogram is built for each video clip and normalized by its $L1$-norm as the feature vector representation. Finally, feature vectors of the training clips are fed into a single-class SVM to model normal crowd behaviors, and this model is tested on the testing clips.

The main differences between our method and the method in [22], which contribute to the improvement of the detection accuracy, lie in three aspects. First, we make more plausible assumptions for the Social Force model, and compute interaction forces from dense trajectories rather than optical flow fields, which enables more accurate estimation of social forces. Second, we extract codewords from interaction force sequences instead of spatio-temporal blocks in force flows, which significantly reduces the computation without loss of detection performance. Third, we use the single-class SVM rather than the LDA model for novelty detection, which is both computationally efficient and accurate.

4. Experiments and Discussion

In this section we evaluate the performance of our method using different features and SVM kernels, and compare with the state-of-the-art method in [22] on two common datasets [1, 2]. Samples of these datasets can be seen in Figure 3.

4.1. Evaluation of features

To gauge how useful each of the possible features (dense trajectories, social force, HOF, HOG, and MBH) are for detecting abnormal crowd behaviors, we first evaluate the performance of each of the features individually. Each feature was tested over 5 trials using the UMN dataset. For each trial the training set was created by randomly selecting 50% of the normal clips. In these tests our SVM is trained using a linear kernel. The receiver operating characteristic (ROC) curves for each feature over all 5 trials are shown in Figure 4. The curves show the true abnormal behavior detection rate versus the false abnormal behavior detection rate. Table

Feature	Average Area and Std. Dev.
Dense Trajectories	0.9088 ± 0.0199
Social Force	0.9743 ± 0.0014
HOF	0.6899 ± 0.0260
HOG	0.6475 ± 0.0234
MBH	0.4922 ± 0.0562

Table 1. Average area under the ROC curve and standard deviations for classification using each feature.

Figure 3. Example videos from the UMN and UCF datasets.

1 shows the means and standard deviations of areas under the ROC curves. These are calculated by averaging the area under the ROC curve for each individual trial.

The results from our trials show the social force feature is the most effective feature for detecting anomalous events with an average area under the ROC curve of 0.97428. Interestingly, dense trajectories alone are quite accurate; for a false detection rate of less than about 2.5% the dense trajectories slightly outperform social force. However, after this the social force feature dominates the other features in ROC space. The results for HOF, HOG and MBH features are underwhelming with areas of 0.68989, 0.64747, and 0.49221, respectively.

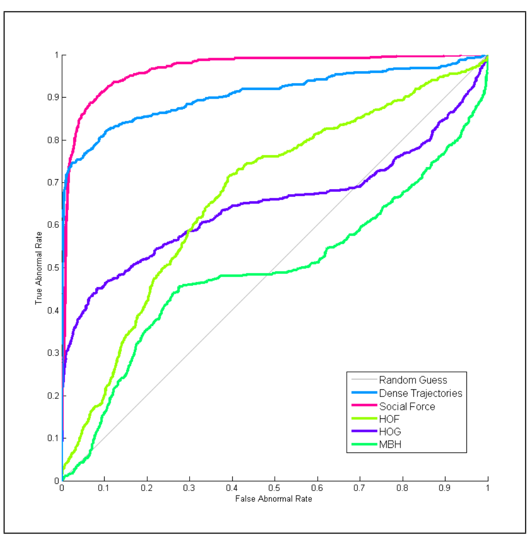

Figure 4. Aggregate ROC curves for classification using each feature.

4.2. Evaluation of kernels

In this section we investigate if it is possible to achieve performance gains by using a different kernel to train our single-class SVM. Potential kernel candidates are a polynomial kernel and a radial basis kernel. For the polynomial kernel we tested the function $k\left(x,y\right) = \left(\gamma\left(x^T y\right) + C\right)^d$ with parameters $C = 0, d = 3, \gamma = \frac{1}{N}$, where N is the number of training features. For the radial basis kernel we tested the function $k\left(x,y\right) = e^{\left(-\gamma|x-y|^2\right)}$ with $\gamma = \frac{1}{N}$, where N is the number of training features. To test these kernels we used the same data from the 5 random trials in the evaluation of features, but changed the kernel for the SVM. The average area under the ROC curves is shown in Figure 5.

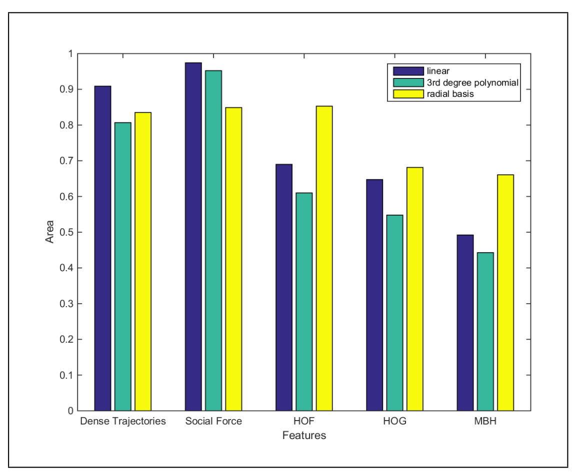

Figure 5. Average area under the ROC curves each feature using a linear, polynomial, and radial basis kernel.

Our results indicate that for dense trajectories and social force the top performing kernel is the linear kernel. However, for HOF, HOG, and MBH features the radial basis kernel drastically improves the performance of the SVM, even exceeding the linear kernel. This is probably explained by more complex decision boundaries for the HOG, HOF, and MBH features compared to dense trajectories and social force. Despite this large performance gain, neither HOF, HOG, or MBH using the radial basis kernel performs better than dense trajectories or social force using a linear kernel.

4.3. UMN dataset

The University of Minnesota (UMN) unusual crowd activity dataset [1] contains 11 videos of 3 different indoor and outdoor scenes. Each video consists of an initial part of normal behaviors and ends with abnormal behaviors of escaping.

In order to compare our performance with [22], we follow the same evaluation procedure as theirs. 5 different videos of the outdoor scene are selected, and only clips with normal behaviors are used for building the codebook and training the SVM. Based on our findings, we use the social force feature alone combined with a linear kernel.

The ROC curves in Figure 6 illustrate that our method outperforms the state-of-the-art method in [22] in detecting

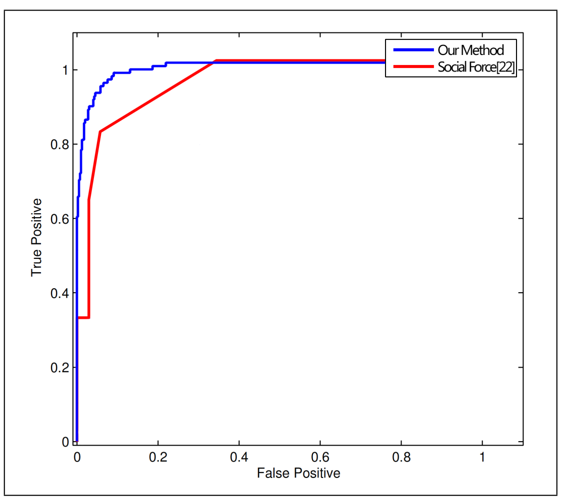

Figure 6. ROC curves for detection of abnormal behaviors in the UMN dataset. Our method (blue) outperforms the state-of-the-art method (red).

Method	Area under ROC
Our Method	0.98
Social Force [22]	0.96

Table 2. Comparison of the use of our method versus the state-of-the-art method for detection of the abnormal behaviors in the UMN dataset.

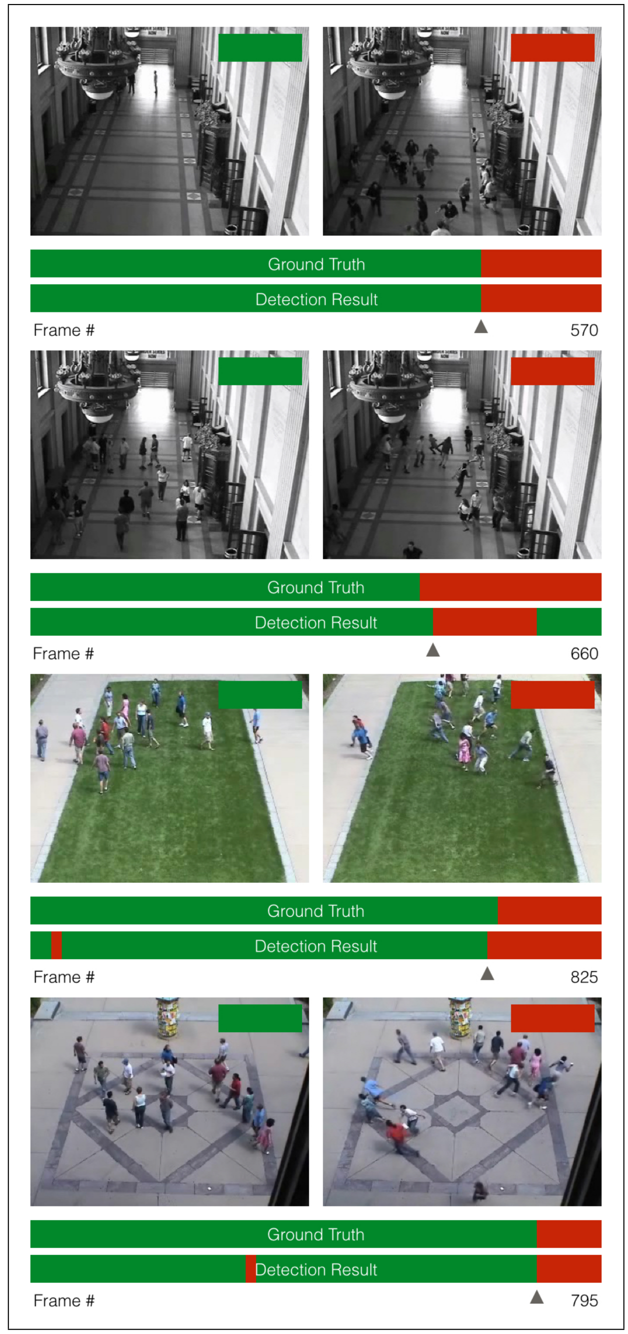

Figure 7. Qualitative results of the abnormal behavior detection for four sample videos of the UMN dataset. Each row shows the results for a video in the dataset. The ground truth bar and the detection result bar indicate the actual and predicted labels of each frame. Green color represents the normal frames, while red represents the abnormal ones. In each row, the left image shows the first frame of the video, and the right image shows a detected abnormal frame. The black triangle on the horizontal bars identify the timing of the shown abnormal frames.

anomalous events, and Table 2 provides the quantitative results of the comparison. If a false detection rate of less than 2.5% is allowed, the true detection rate is 86.84%, and the overall detection accuracy is 95.29%.

Figure 7 shows some of the qualitative results for detection of abnormal scenes. Each row in the figure illustrates our detection results for a video in the dataset. Only the first of the four sampled videos is in the training set, while the others are not. The results show that our method achieves a very high accuracy on the training videos, and does a decent job on the other videos as well.

4.4. UCF web dataset

The University of Central Florida (UCF) web dataset [2] contains more challenging videos of real-world crowds that have been collected from various online sources. It contains 20 videos which are divided into 12 normal crowd scenes such as pedestrian walking and marathon running, and 8 abnormal scenes of escape panics, protesters clashing and crowd fighting.

In order to compare our performance with [22], we follow the same evaluation procedure as theirs. We randomly exclude 2 videos from the normal set and train on the rest. In the testing phase we add the excluded videos to the test set. We run this experiment 10 times and plot the ROC curve by averaging the results of these experiments. Based on our findings, we use the social force feature alone combined with a linear kernel.

The ROC curves in Figure 8 illustrate that our method outperforms the state-of-the-art method in [22] in detecting anomalous events, and Table 3 provides the quantitative results of the comparison. If a false detection rate of less than 2.5% is allowed, the true detection rate is 66.15%, and the overall detection accuracy is 88.22%.

Figure 9 shows some of the qualitative results for detection of abnormal scenes. Although our method is able to classify many of the scenes correctly, it has some obvious weaknesses. For example, it cannot really distinguish racing from escaping. The scenes in the UCF web dataset are very diverse, and people in the scenes come with different sizes and perspectives, which makes the anomaly detection task much more difficult.

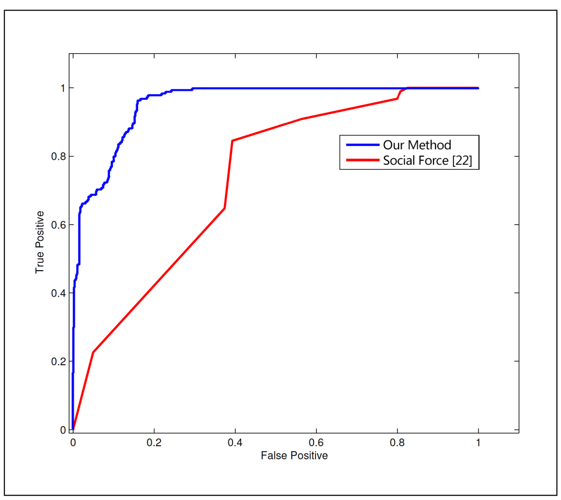

Figure 8. ROC curves for detection of abnormal behaviors in the UCF web dataset. Our method (Blue) outperforms the state-of-the-art method (Red).

Method	Area under ROC
Our Method	0.96
Social Force [22]	0.73

Table 3. The comparison of the use of our method and the state-of-the-art method for detection of the abnormal behaviors in the UCF web dataset.

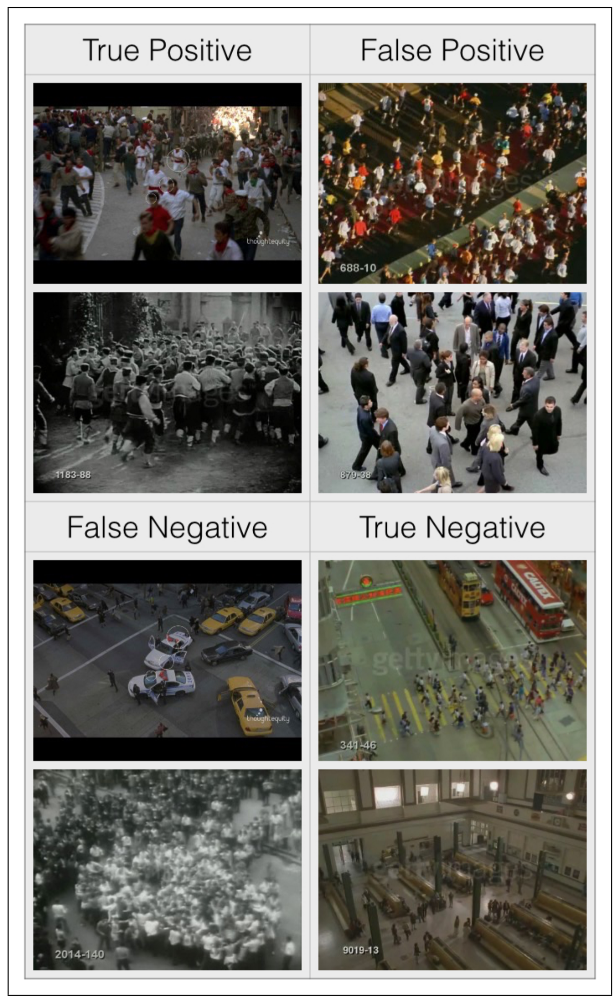

Figure 9. Qualitative results of the abnormal behavior detection for videos of the UCF web dataset. Two sampled frames that fall in each cell are showed here.

5. Conclusion

As the need for electronic video surveillance continues to rise in the future it will be of increasing necessity to be able to prescreen video feeds for possible abnormal behaviors. Advance warning of such behaviors can reduce the time needed to engage emergency services and could ultimately save lives. In this work we present a method for anomaly detection using dense trajectories to compute social interactive forces and then detect abnormalities in interactive forces using a single-class SVM. Our approach outperforms prior works on both the UMN and UCF datasets when using social force as the training features.

Our analysis was conducted entirely post-hoc. We would like to explore the possibility of using online detections in streaming video. This advancement is critical to making abnormal behavior detection common-place. Online detection should be possible if a codebook is computed in advance as codebook computation (using K-means) is the most time consuming element of the system. We have also not explored localization of abnormal events within a clip. In future work we could explore localization as well as many other factors including optimizing the codebook size, combining different features for classification, and voting using individual features. Additionally, other classification techniques could be explored such as artificial neural networks (ANN).

Acknowledgements

We would like to briefly thank Dr. Vikas Singh and Jia Xu. Their feedback and instruction ultimately led to the creation of this work.

References

[1] Detection of unusual crowd activity dataset. http://mha.cs.umn.edu/proj_events.shtml#crowd. 4, 6

[2] Ucf web videos dataset. http://www.vision.eecs.ucf.edu/projects/rmehran/cvpr2009/Abnormal_Crowd.html#VI._Downloads. 4, 7

[3] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz. Robust real-time unusual event detection using multiple fixed-location monitors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(3):555–560, 2008. 1, 2

[4] A. Basharat, A. Gritai, and M. Shah. Learning object motion patterns for anomaly detection and improved object detection. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008. 1, 2

[5] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003. 2, 3

[6] O. Boiman and M. Irani. Detecting irregularities in images and in video. *International Journal of Computer Vision*, 74(1):17–31, 2007. 1, 2

[7] G. J. Brostow and R. Cipolla. Unsupervised bayesian detection of independent motion in crowds. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 594–601. IEEE, 2006. 1, 2

[8] Y. Cong, J. Yuan, and J. Liu. Sparse reconstruction cost for abnormal event detection. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3449–3456. IEEE, 2011. 1, 2

[9] X. Cui, Q. Liu, M. Gao, and D. N. Metaxas. Abnormal detection using interaction energy potentials. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3161–3167. IEEE, 2011. 2

[10] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005. 2

[11] N. Dalal, B. Triggs, and C. Schmid. Human detection using oriented histograms of flow and appearance. In *Computer Vision–ECCV 2006*, pages 428–441. Springer, 2006. 2

[12] G. Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Proceedings of the 13th Scandinavian Conference on Image Analysis*, LNCS 2749, pages 363–370, Gothenburg, Sweden, June-July 2003. 2

[13] D. Helbing, I. Farkas, and T. Vicsek. Simulating dynamical features of escape panic. *Nature*, 407(6803):487–490, 2000. 2

[14] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282, 1995. 2

[15] L. Itti and P. Baldi. A principled approach to detecting surprising events in video. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 631–637. IEEE, 2005. 1, 2

[16] J. Kim and K. Grauman. Observe locally, infer globally: a space-time mrf for detecting abnormal activities with incremental updates. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2921–2928. IEEE, 2009. 1, 2

[17] L. Kratz and K. Nishino. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1446–1453. IEEE, 2009. 1, 2

[18] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008. 2

[19] W. Li, V. Mahadevan, and N. Vasconcelos. Anomaly detection and localization in crowded scenes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(1):18–32, 2014. 1, 2

[20] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos. Anomaly detection in crowded scenes. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1975–1981. IEEE, 2010. 1, 2

[21] J. S. Marques, P. M. Jorge, A. J. Abrantes, and J. Lemos. Tracking groups of pedestrians in video sequences. In *Computer Vision and Pattern Recognition Workshop, 2003.*

CVPRW'03. Conference on, volume 9, pages 101–101. IEEE, 2003. 1, 2

[22] R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 935–942. IEEE, 2009. 1, 2, 3, 4, 6, 7

[23] R. Raghavendra, A. Del Bue, M. Cristani, and V. Murino. Abnormal crowd behavior detection by social force optimization. In *Human Behavior Understanding*, pages 134–145. Springer, 2011. 2

[24] V. Saligrama and Z. Chen. Video anomaly detection based on local statistical aggregates. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2112–2119. IEEE, 2012. 1, 2

[25] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt. Support vector method for novelty detection. In *NIPS*, volume 12, pages 582–588, 1999. 1, 3

[26] N. T. Siebel and S. Maybank. Fusion of multiple tracking algorithms for robust people tracking. In *Computer VisionECCV 2002*, pages 373–387. Springer, 2002. 1, 2

[27] P. Tu, T. Sebastian, G. Doretto, N. Krahnstoever, J. Rittscher, and T. Yu. Unified crowd segmentation. In *Computer Vision–ECCV 2008*, pages 691–704. Springer, 2008. 1, 2

[28] B. Wang, M. Ye, X. Li, F. Zhao, and J. Ding. Abnormal crowd behavior detection using high-frequency and spatio-temporal features. *Machine Vision and Applications*, 23(3):501–511, 2012. 1, 2

[29] H. Wang, A. Klaser, C. Schmid, and C.-L. Liu. Action recognition by dense trajectories. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3169–3176. IEEE, 2011. 1, 2, 3

[30] H. Wang, A. Kläser, C. Schmid, and C.-L. Liu. Dense trajectories and motion boundary descriptors for action recognition. *International Journal of Computer Vision*, 103(1):60–79, 2013. 1, 2, 3

[31] H. Wang and C. Schmid. Action recognition with improved trajectories. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 3551–3558. IEEE, 2013. 1, 2

[32] S. Wu, B. E. Moore, and M. Shah. Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2054–2060. IEEE, 2010. 1, 2

[33] T. Zhang, H. Lu, and S. Z. Li. Learning semantic scene models by object classification and trajectory clustering. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1940–1947. IEEE, 2009. 1, 2

[34] Y. Zhang, L. Qin, H. Yao, and Q. Huang. Abnormal crowd behavior detection based on social attribute-aware force model. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 2689–2692. IEEE, 2012. 2