Dynamic Texture Reconstruction from Sparse Codes for Unusual Event Detection in Crowded Scenes

Jingxin Xu[*] , Simon Denman[*] , Sridha Sridharan[*] ,Clinton Fookes[*] ,Rajib Rana[†]
j15.xu@qut.edu.au, s.denman@qut.edu.au, s.sridharan@qut.edu.au,
c.fookes@qut.edu.au, rajeevrans@gmail.com

[*]Image and Video Laboratory
Queensland University of Technology
GPO Box 2434, Brisbane 4001, Australia

CSIRO ICT Centre
Australia

ABSTRACT

Unusual event detection in crowded scenes remains challenging because of the diversity of events and noise. In this paper, we present a novel approach for unusual event detection via sparse reconstruction of dynamic textures over an overcomplete basis set, with the dynamic texture described by local binary patterns from three orthogonal planes (LBP-TOP). The overcomplete basis set is learnt from the training data where only the normal items observed. In the detection process, given a new observation, we compute the sparse coefficients using the Dantzig Selector algorithm which was proposed in the literature of compressed sensing. Then the reconstruction errors are computed, based on which we detect the abnormal items. Our application can be used to detect both local and global abnormal events. We evaluate our algorithm on UCSD Abnormality Datasets for local anomaly detection, which is shown to outperform current state-of-the-art approaches, and we also get promising results for rapid escape detection using the PETS2009 dataset.

Categories and Subject Descriptors

I.4 [**Image Processing and Computer Vision**]: Scene Analysis

General Terms

Algorithms

Keywords

Sparse Coding, Anomaly Detection, Dynamic Texture, Dantzig Selector, Compressed sensing

1. INTRODUCTION

One goal of intelligent surveillance is to apply computer vision and machine learning techniques to detect unusual events in crowded scenes [6]. In order to investigate this problem, we first need to address two points: the definition of unusual events, and the features to represent the events.

Conventionally the unusual events are identified as those events which occur with a low probability [7, 11, 18] because of the following two reasons: 1) it is usually very difficult to list all events that can possibly occur in a surveillance environment due to the diversity of the events; and 2) the usual events are more likely to occur compared to the unusual ones. As a result, unusual event detection becomes a novelty detection problem.

Early applications of event detection extracted the trajectories of moving objects as features and heavily relied on object tracking [8]. However, due to the clustering of people in groups and occlusions, object tracking is not reliable in densely crowded scenes. Current state-of-the-art algorithms [1, 7, 11, 18] focus on extracting local motion features. Previously, many algorithms extract local motion features based on computing optical flow [1, 7, 18]. However, optical flow is often unreliable in textureless regions [17]. In addition, besides the motion features, appearance features are also useful for unusual event detection [11], while visual information obtained from optical flow is very limited [17]. Dynamic textures are sequences of images of movement that exhibit spatio-temporal stationary properties [5]. Recent research [11] have been shown that dynamic texture is more suitable for unusual event detection in crowded scenes than optical flow. However, the dynamic textures are represented in Autoregressive moving averaging (ARMA) model in [11], while research in facial expression recognition [21] indicates that the descriptor as the histogram of Local Binary Patterns from Three Orthogonal Planes (LBP-TOP) is more suitable to represent dynamic texture.

In the recent years, sparse representation [10, 14, 19, 20, 22] has become popular in computer vision. One of the most widely used technique in this field is sparse coding [10, 14], which is a generative model motivated by the spatial receptive fields in the visual cortex. Ranzato et al.[12] summarise the benefits of this sparse overcomplete representation: 1) data that is non-linearly separable is more likely to be linearly separable in a higher dimensional space (overcomplete); 2) simple interpretation of the input patterns (sparse); and 3) consistent with the biological vision processes. Yang et al.[19] use sparse coding to replace K-means to learn a codebook in the joint “bags of features” and spatial pyramid framework. The benefit of sparse coding in this step is that in the classification stage, one can use a linear SVM to achieve the performance of a non-linear SVM with the codebook generated through K-means. Very recently, [20] presents a novel algorithm for abnormal event detection based on the sparse reconstruction cost for multi-level histogram of optical flows.

This sparse representation is also critical in compressive sensing [2–4], which has also been applied in computer vision [22]. Since both sparse coding and compressive sensing adopt the same sparse representation, [16] combines them together in speech recognition to achieve an improvement over state-of-the-art algorithms.

1.1 Overview of the Proposed Algorithm

In this paper, we present a novel algorithm for unusual event detection in crowded scenes based on novelty detection from dynamic textures (LBP-TOP). In our application, principle component analysis (PCA) is used for dimension reduction and whitening of the LBP-TOP dynamic textures. We represent the whitened dynamic textures as a linear superposition of sparse coefficients on an overcomplete basis set, which are learnt from the normal items in the training process using sparse coding [10, 14]. Traditionally sparse coding models a local image patch [14]. In our application, this model is extended to spatio-temporal patches to model dynamic textures. Let y be the motion pattern (dynamic texture), B be the basis functions, and x be the sparse coefficients. Using this notation, we have the following relationship,

$$y = Bx, \tag{1}$$

where $y \in R^N$, $x \in R^M$ and $B \in R^{N \times M}$. Here the coefficient x is sparse, which means most of the entries are 0 and the basis set is overcomplete, which means that the number of basis functions is larger than the dimension of the input ($M > N$). Learning the basis functions is achieved by minimising the below objective function,

$$f(x) = \parallel y - Bx \parallel_2^2 + \lambda \parallel x \parallel_1, \tag{2}$$

where λ is a constant, $\parallel y - Bx \parallel_2$ is the reconstruction error, and the $\parallel x \parallel_1$ is the penalty function. Here we use the l_1 norm as the penalty function.

In the detection process, given the learnt overcomplete basis set B and an new observation y, the sparse coefficient x is computed by the Dantzig Selector [3], which is an algorithm for ℓ_1 norm minimization problem and has been used in compressive sensing [4]. Let $\hat{y} = Bx$ be denoted as the reconstructed signal. The Euclidean distance between $\hat{y}$ and y is is the reconstruction error, which is used as the criterion for unusual event detection. That is because, the basis functions are learnt from the training data where only the normal items observed, and the abnormal items would expect to have high reconstruction error over the normal basis set.

Our application is divided into two parts: local anomaly detection and global anomaly detection. Figure 1 illustrates the architecture of our application for local anomaly detection. We divide the video scene into regular spatio-temporal patches from which the dynamic textures (LBP-TOP) are extracted. As a result, the local anomaly detection is able to generate alarms at abnormal locations. We show empirical results on the UCSD datasets. The global anomaly event is defined as the combination of co-existing local dynamic textures over the whole scene. We concatenate all histogram in each spatio-temporal patches on the entire scene, and detect global abnormal events (rapid escape) using similar approach. We show promising results on the PETS 2009 dataset.

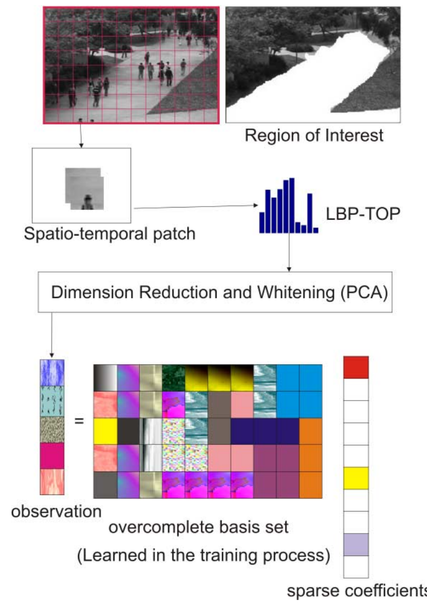

Figure 1: Overview of our application: 1) the video scene is divided into spatio-temporal patches; 2) for those patches in the region of interest, LBP-TOP features are extracted; 3) PCA is used for whitening and dimension reduction; 4) an input pattern is represented as a sparse linear superposition over an over complete basis set; 5) in the training process, the overcomplete basis set is learnt based on the normal observations; 6) in the detection process, given the basis set learnt in the training procedure and the input observation, the sparse coefficients are computed and the reconstruction error is defined; 7) the unusual events are identified as those dynamic textures with high reconstruction error.

1.2 Connection to existing work

Unusual event detection in crowded scenes is very challenging because of the diversity of events and the noise in the scenes. Adam et al. [1] use Gaussian Mixture Model (GMM) for unusual event detection. However, their application can only detect local abnormal events. Meanwhile, if the dimension of the feature vector in a GMM system is large, there is often an overfitting problem as the covariance matrix approaches to singular. Wang et al. [18] and Hospedales et al. [7] apply topic models to activity modelling. Their algorithms can detect both local and global abnormal events. However, algorithms using topic models do not easily to support online learning. Cong et al. [20] applied sparse coding to event detection. Their application is able to detect both local and global abnormal events, and support online learning. They show that the sparse representation (See Section 1) outperforms state-of-the art algorithms in several public datasets. However, compared to [20], our approach has the following advantages:

1. In [20], given a new observation y and the learnt basis B, they compute the sparse coefficients based on minizing $\| y - Bx \|_2^2 + \lambda \| x \|_1$. The minimum is called sparse reconstruction cost (SRC), and the unusual events are detected on a threshold of SRC [1]. In our approach, we compute the sparse coefficients based on minimizing $\| x \|_1$, subject to $\| B^T(Bx - y) \|_\infty \leqq \epsilon$, using the Dantzig Selector [3]. Based on this optimal x, we compute the reconstructed signal $\hat{y} = Bx$, and detect the abnormal events by thresholding $\| y - \hat{y} \|_2$. This is a more sensible criterion and we discuss this in detail in Section 2.4. To the best of the author's knowledge, this is the first attempt to use the Dantzig Selector, which is proposed in compressed sensing literature, for unusual event detection.

2. LBP-TOP can capture more appearance features than optical flow, which improves performance. To the best of our knowledge, this is the first attempt to apply sparse coding to LBP-TOP.

It should be noted that, there is some literature in facial expression recognition applies compressed sensing to the traditional LBP descriptor [22]. However, [22] does not use sparse coding. The difference between sparse coding and compressed sensing is obvious. Compressed sensing is proposed in signal processing to reconstruct a signal from incomplete frequencies. As a result, the basis is Fourier bases, and there is no need to learn the basis set in the training process.

2. SPARSE RECONSTRUCTION FOR UNUSUAL EVENT DETECTION

In this section, we illustrate our algorithm in detail. The following five steps are the key elements in our algorithm: 1) LBP-TOP extraction from spatio-temporal patches; 2) Dimension reduction and whitening using PCA; 3) Learning the overcomplete basis set using an efficient sparse coding algorithm; 4) Sparse reconstruction from learnt basis functions using Dantzig Selector from compressed sensing; and 5) Global anomaly detection.

Section 2.1, 2.2, 2.3, 2.4 and 2.5 address each of the above steps respectively.

2.1 LBP-TOP based dynamic texture

In [13], the texture, T, is defined as the joint distribution of intensities from the nine pixels in a 3×3 neighbourhood:

$$T = p(g_0, g_1, g_2, g_3, g_4, g_5, g_6, g_7, g_8), \tag{3}$$

where $g_i (i = 0, ..., 8)$ are the intensities of the pixels, and g_2,g_4,g_6,g_8 are computed by interpolation. Then [13] defines the gray scale invariant local binary pattern (LBP) by considering only the sign of the differences,

$$LBP_8 = \sum_{i=1}^{8} s(g_i - g_0) 2^{i-1}, \tag{4}$$

[1]There is some extra detail in this procedure not given here, however this is not the key point in our discussion

where,

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}. \tag{5}$$

This representation of LBP can be further extended to support rotational invariance. However, as our target application uses a single stationary camera, we don't require this extension. Over a local region (typically much larger than 3×3), the histogram of LBPs can be used to represent the texture.

Dynamic texture extends the traditional spatial texture into the temporal domain. Correspondingly, [21] extends the LBP into a spatio-temporal volume to model dynamic textures. Let $P(x_c, y_c, t_c)$ be the centre pixel in a spatio-temporal neighbourhood. The volume LBP (VLBP) is defined as the joint distribution of the intensities of $3 \times P + 3$ pixels on the current frame, t_c, the previous frame, $t_c - L$, and the next frame, $t_c + L$ in,

$$VLBP(x_c, y_c, t_c) = \sum_{q=0}^{3P+1} s(g_q - g_c) \times 2^q, \tag{6}$$

where P is the number of neighbours in each frame, L is the temporal interval, g_q is neighbour pixels' intensities, and g_c is the centre pixel intensity.

In order to reduce the total number of patterns, [21] further simplifies this model, only calculating the local binary patterns from three orthogonal planes (LBP-TOP) . LBPs are computed with the histogram of the output in each plane. Then the three histograms are concatenated into a single histogram.

For the application of anomalous event detection, we partition the scene into spatio-temporal patches. Within each patch, LBP-TOP is extracted. In each plane we use the 8 pixel neighbourhood. As a result, each plane contains 2^8 local binary patterns. Among the three planes, XY contains rich appearance features. XT and YT contains the motion features with limited appearance features. Similar to [9], only the XT and YT are considered in our application to make it robust to human appearance. The size of the histogram in our application is $2^8 \times 2 = 512$ bins.

2.2 Dimension Reduction and Whitening

In this section, we discuss a preprocessing step for the LBP-TOP features extracted in Section 2.1, which includes dimension reduction and whitening. We can achieve the two tasks using PCA. The reason to do dimension reduction includes both: 1) computational simplicification; and 2) avoid overfitting. The reason for the whitening process is linked to the objective function in Equation (2), where both the reconstructed signal $\hat{y} = Bx$ and original signal y are N dimensional vectors. An accurate measure of the distance between two vectors should be the Mahalanobis distance. However, in sparse coding, the reconstruction error between $\hat{y}$ and y is computed as the Euclidean distance. The whitening process transforms the data into a new space, so that the covariance matrix in the transformed space becomes an identity matrix. In this case, the Euclidean distance is equavalent to the Mahalanobis distance [17].

To simplify the problem, we center the mean at origin. Then the data is transformed to their principal components in the following method [15]: 1) compute the covariance matrix for the full training dataset; 2) the eigenvalues and eigenvectors are computed and sorted by decreasing the eigenvalues; 3) the K eigenvectors with the largest K eigenvalues form a matrix A, with each column an eigenvector; and 4) let e be the data, and $\hat{e}$ be the principal component, then we have,

$$\hat{e} = A^T e, \tag{7}$$

where A^T is the transpose matrix of A.

The covariance matrix of the principal components are a diagonal matrix. The last step in the whitening process is to normalise the variance. Let,

$$\Lambda = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{\lambda_k}} \end{bmatrix}, \tag{8}$$

be a $K \times K$ diagonal matrix where $\lambda_1 \cdots \lambda_k$ represents the K largest eigenvalues sorted in decreasing order. Thus the final transform in the preprocessing step is,

$$\hat{e} = \Lambda A^T e. \tag{9}$$

In our application, we reduce the dimensionality from 512 into 256, thus $K = 256$ in our application.

2.3 Learning the overcomplete basis set

In the training process, we learn the overcomplete basis set, which is the following optimization problem

$$\{B, x\} = arg\,min \left\{ \parallel y - Bx \parallel_2^2 + \lambda \parallel x \parallel_1 \right\}. \tag{10}$$

We make use of the efficient sparse coding algorithm provided in [10]. The design of this algorithm is such that, when B is fixed, the objective function is convex in x; and when x is fixed, the objective function is convex in B. As a result, we first fix B, and optimize x; then fix x, and optimize B. This process is conducted iteratively until the algorithm converges.

When we fix B, the optimization of the sparse coefficients, x, relies on the feature sign search algorithm [10].

In the next stage we fix x, and optimise the basis functions B, using the Lagrange Dual algorithm in [10].

2.4 Sparse reconstruction using Dantzig Selector

The basis set B is learnt in the training process as described in Section 2.3, where only normal items are observed. In the detection process, we need to compute the reconstruction error for a new observation.

Given a new observation y, we need to compute the sparse coefficients x using B, which has been learned in the training process. In this section, we introduce the Dantzig Selector algorithm [3] of compressed sensing, to compute the sparse coefficients. Once x has been computed, the reconstruction error is defined as $\parallel y - Bx \parallel_2$.

Compressive sensing addresses the problem of signal reconstruction from highly incomplete frequencies [4]. Let $K \in R^{M'}$ be the discrete Fourier coefficients of a discrete signal sampled at the Nyquist Rate. Let $V \in R^{N'}$ be a subset of K, where $M' > N'$. In terms of the theory in compressive sensing [2], we have,

$$V = \Phi K = \Phi \Psi s = \Theta s, \tag{11}$$

where $\Phi \in R^{N' \times M'}$, $\Psi \in R^{M' \times M'}$, $s \in R^{M'}$, $\Theta \in R^{N' \times M'}$ and s is the sparse coefficients. In the reconstruction part, given the compressed signal V, we need to compute s in order to reconstruct K. We need to solve the following optimization problem,

$$\hat{s} = arg\,min \parallel s \parallel_1 \quad s.t.\ V = \Theta s. \tag{12}$$

Please note that, if we let $B = \Theta$, $x = s$, $y = V$, $M = M'$ and $N = N'$, we can compute the sparse coefficients x,

$$\hat{x} = arg\,min \parallel x \parallel_1 \quad s.t.\ y = Bx. \tag{13}$$

In Section 2.3, we have introduced that in the learning process of sparse coding, we can fix B, and compute the coefficient by minmising $\parallel y - Bx \parallel_2^2 + \lambda \parallel x \parallel_1$. In [20], this approach is used to compute the sparse coefficient in the detection process. Please notice that, here λ is a fixed value which is set manually. The drawback of this approach is that, λ might not be properly set, and the performance depends on this parameter setting. The recovery algorithm of compressive sensing minimize the ℓ_1 norm while setting $y = Bs$ as a constraint. There is no such λ. The Dantzig Selector [3] relaxes the equality constraint through residual bounded correlation, by minimizing $\parallel x \parallel_1$, subject to $\parallel B^T(Bx - y) \parallel_\infty \leqq \epsilon$, where ϵ is a small value. In our experiments, we use the feature sign search algorithm (See Section 2.3) and the Dantzig Selector [2] to compute the sparse coefficients and we compare their performance. We show that the Dantzig Selector has better performances.

Because the overcomplete basis set is learnt from normal observations in the training process, the abnormal observations will cause large reconstruction errors. We detect the unusual events based on a threshold on the reconstruction errors.

2.5 Global Anomaly Detection

In previous sections, we have described our algorithm for local anomaly detection. This section will introduce global anomaly detection. Here global anomaly means the combination of coexisting patterns in the whole scene. Typically, we can build an application based on concatenating the histograms in each local spatio-temporal patches together. Then the same learning and detection process is used. However, before we do that, we need to remove the influence of the background. We compute the means of all the bins along the training data. This forms the background model. Then for each input histogram, we subtract the background component. In this way, the bins related to the background become very small, while patterns caused by motions are pronounced.

3. EXPERIMENTS

[2] available at http://users.ece.gatech.edu/ justin/l1magic/

Figure 2: The detected unusual events in Peds1.The unusual events are marked in red. They are also enclosed in green to improve clarity.

	Adam et al. [11]	MDT [11]	Sparse	Sparse-CS
Ped1	38%	25%	33.10%	31.03%
Ped2	42%	25%	8.6%	5.65%
average	40%	25%	20.85%	18.34%

	Adam et al.[11]	MDT [11]	SRC [20]	Sparse	Sparse-CS
Ped1	24%	45%	46%	53.76%	57.43%

Table 1: Equal Error Rates for Frame based ground truth (Top). Detection Rates for the Location based ground truth(Bottom). Here "sparse" represents the approach of using the feature sign search algorithm to compute the sparse coefficient. The "sparse+cs" is our proposed approach using the Dantzig Selector to compute the sparse coefficients.

3.1 Local Anomaly Detection

We use the UCSD Abnormality Dataset[3] [11] for evaluation. The UCSD datasets contain video clips of two pedestrian scenes from a campus, Peds1 and Peds2. The Peds1 dataset contains clips of groups of people walking towards and away from the camera, with a resolution of 158×238 (34 clips for training, and 36 clips for testing). The Peds2 dataset contains scenes of pedestrians moving parallel to the camera plane, with a resolution of 360×240 (16 clips for training, and 12 clips for testing). All clips are approximately 200 frames long. Examples of anomalies include a bus, a wheelchair, a bicycle, and a skater, and these abnormal events only exist in the test data. The UCSD dataset contains frame-level ground truth and pixel-level ground truth. In order to compare algorithms, we use the evaluation method presented in [11] for pixel level detection.

We divide the scene into $20 \times 20 \times 11$ size spatio-temporal patches. We set the λ in Equation (10) as 0.2. The number of basis functions is 512, which is larger than our dimensionality (256, see Section 2.2). We show the results in Figure (3) and Table 1. We show that our approach outperforms several state of the art algorithms. Meanwhile, we show that if we use the Dantzig Selector to compute the sparse coefficients, we can have better performance than using the feature sign search algorithm [10].

3.2 Global Anomaly Detection

For global anomaly detection, we use the PETS 2009 dataset[4]. Here we aim to detect the rapid escape. We use the regular flow data as the training data set, and the Time14-33 in S3.HL dataset as the test data. There are four camera views, which are View001, View002, View003, and View004. In the test sequences, each view has 378 frames. The rapid escape begins around Frame 341. Thus we manually annotate the frames after Frame 340 as abnormal. Based on this ground truth, we compute the ROC curve and AUC. We resize the images into 192×144 resolution, and convert the colour images into gray level, and the patch size is $48 \times 48 \times 5$. Figure (4) illustrates the normal and abnormal scene in View 001. Figure (5) shows the ROC curves, and Table 2 shows the areas under the ROC curves. It can be seen that the proposed approach is able to detect global anomalies successfully.

[3]http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm
[4]http://www.cvg.rdg.ac.uk/PETS2009/

Figure 4: View 001 Data. The left image is a normal frame. The right image is an abnormal frame.

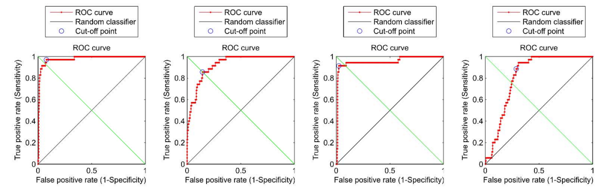

Figure 5: ROC curves for global anomaly detection. From left to right, we shows results for View001, View002, View003, View004 respectively

View	AUC
view001	0.97726
view002	0.92468
view003	0.95266
view004	0.81860

Table 2: AUC for global anomaly detection

4. CONCLUSION

This paper presents a novel approach for unusual event detection based on the sparse reconstruction error over an overcomplete basis set learnt from the LBP-TOP based dynamic textures, where only the normal events exist in the training dataset. We show that if we use the Dantzig Selector algorithm to compute the sparse coefficients in the reconstruction section, performance can be improved significantly. We demonstrate that this approach can be applied to both local and global anomaly detection successfully.

5. REFERENCES

[1] 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* vol. 30, no. 3, pp. 555–560, 2008.

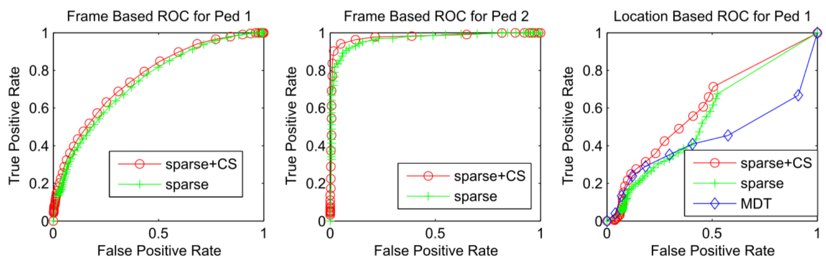

Figure 3: ROC curves of experimental results. The red line represents the ROC curves using the Dantzig Selector, which is an algorithm from compressive sensing to compute the sparse coefficients("sparse+cs"). The green line represents the ROC curves using feature sign search algorithm to compute the sparse coefficient ("sparse"). The left and middle images are the frame-based ROC curves for Peds1 and Peds2 respectively. The right image is the location-based ROC curves for Peds1. The blue line represents the ROC curve of MDT[11].

[2] R. Baraniuk, "Compressive sensing," *Lecture notes in IEEE Signal Processing magazine*, vol. 24, no. 4, pp. 118–120, 2007.

[3] E. Candes and T. Tao, "The dantzig selector: statistical estimation when p is much larger than n," *The Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 2007.

[4] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489 – 509, feb. 2006.

[5] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto, "Dynamic textures," *International Journal of Computer Vision*, vol. 51, pp. 91–109, 2003.

[6] C. Fookes, S. Denman, R. Lakemond, D. Ryan, S. Sridharan, and M. Piccardi, "Semi-supervised intelligent surveillance system for secure environments," in *Industrial Electronics (ISIE), 2010 IEEE International Symposium on*, 2010, pp. 2815 –2820.

[7] T. Hospedales, S. Gong, and T. Xiang, "A markov clustering topic model for mining behaviour in video," in *Proc. IEEE 12th Int Computer Vision Conf*, 2009, pp. 1165–1172.

[8] N. Johnson and D. Hogg, "Learning the distribution of object trajectories for event recognition," *Image and Vision Computing*, vol. 14, no. 8, pp. 609 – 615, 1996, 6th British Machine Vision Conference. [Online]. Available:

[9] V. Kellokumpu, G. Zhao, and M. Pietikäinen, "Human activity recognition using a dynamic texture based method," in *In BMVC*, 2008.

[10] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *In NIPS*. NIPS, 2007, pp. 801–808.

[11] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 1975–1981.

[12] C. Marc Aurelio Ranzato, S. Chopra, and Y. LeCun, "Efficient learning of sparse representations with an energy-based model," *Advances in neural information processing systems*, vol. 19, 2006.

[13] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971 –987, Jul. 2002.

[14] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *Vision Research*, vol. 37, no. 23, pp. 3311 – 3325, 1997. [Online]. Available:

[15] D. G. Richard O.Duda, Peter E.Hart, *Pattern Classification*. John Wiley & Sons, 2001.

[16] G. Sivaram, S. Nemala, M. Elhilali, T. Tran, and H. Hermansky, "Sparse coding for speech recognition," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, march 2010, pp. 4346 –4349.

[17] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer, 2011.

[18] X. Wang, X. Ma, and W. E. L. Grimson, "Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models,"*Pattern Analysis and Machine Intelligence, IEEE Transactions on* vol. 31, no. 3, pp. 539–555, 2009.

[19] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, june 2009, pp. 1794 –1801.

[20] J. L. Yang Cong, Junsong Yuan, "Sparse reconstruction cost for abnormal event detection," in *CVPR 2011*, 2011.

[21] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 6, pp. 915 –928, 2007.

[22] M.-W. Huang, Z. wei Wang, and Z.-L. Ying, "A new method for facial expression recognition based on sparse representation plus lbp," in *Image and Signal Processing (CISP), 2010 3rd International Congress on*, vol. 4, oct. 2010, pp. 1750 –1754.