International Journal of Communication Networks and Information Security (IJCNIS) Vol. 4, No. 1, April 2012

Automated Intruder Detection from Image Sequences using Minimum Volume Sets

Tarem Ahmed[1,2], Xianglin Wei[3], Supriyo Ahmed[2] and Al-Sakib Khan Pathan[1]

[1]Department of Computer Science, International Islamic University Malaysia, Kuala Lumpur, Malaysia
[2]Department of Electrical and Electronic Engineering, BRAC University, Dhaka, Bangladesh
[3]Department of Computer Science and Engineering, PLA University of Science and Technology, Nanjing, China
{tarem, wei_xianglin}@ieee.org, supriyo@bracu.ac.bd, sakib@iium.edu.my

Abstract**:** We propose a new algorithm based on machine learning techniques for automatic intruder detection in visual surveillance networks. The proposed algorithm is theoretically founded on the concept of Minimum Volume Sets. Through application to image sequences from two different scenarios and comparison with existing algorithms, we show that it is possible for our proposed algorithm to easily obtain high detection accuracy with low false alarm rates.

Keywords**:** automated surveillance, online anomaly detection, real-time outlier detection, learning algorithms**.**

1. Introduction

An extensive network of multimodal surveillance systems is prevalent in many places in today's world. The London Underground and London Heathrow airport have more than 5000 cameras, for instance [1]. Simultaneously monitoring multiple image trains becomes tedious and monotonous for human operators with typically short attention spans and cognitive limits on how many screens one may attentively observe simultaneously. The goals of current research in autonomous surveillance are to develop algorithms that attract the attention of a human operator in real-time based on end-user requirements, process information arriving from a multi-sensor environment at high rates, and use inexpensive, standard components [1, 2].

We propose the One-Class Neighbor Machine (OCNM) algorithm [3], run using a sliding window implementation, to autonomously detect, in real-time, the occurrence of an anomalous image in a sequence of images being captured by a visual surveillance network. Through application to two different and complementary scenarios, and comparisons with representative algorithms from two families of algorithms commonly used for novelty detection in image sequences, we demonstrate that our proposed algorithm not only achieves superior performance, but also provides near-perfect detection accuracy with low false alarm rates.

The applicability of the OCNM algorithm to this problem had previously been shown in [4], through experiments conducted on a sequence of images collected from an example Closed-Circuit Television (CCTV) surveillance system. This paper, however, presents a more extensive treatment of the problem. First, the size of the real-world CCTV data set has been increased. Second, all experiments are repeated on a specially-setup indoor network of higher-resolution cameras, using a distributed monitoring architecture. This is complementary to the CCTV data set, which employs a centralized monitoring architecture in an outdoor environment. Furthermore, an additional data pre-processing block, namely the Canny edge detector [5], is employed to filter out unnecessary illumination and hue information from the better-quality indoor images. Finally, discussions on the algorithm computational complexities have been added.

This paper is organized as follows. We motivate and describe our proposed OCNM algorithm in Section 2. Section 3 describes two related algorithms that we compare our proposed algorithm against. Section 4 presents experimental results on real footage from an example, simple CCTV surveillance system that is deployed in an outdoor environment and uses centralized monitoring architecture. Section 5 presents experimental results on a network of cameras setup in an indoor environment that uses a distributed monitoring architecture. Section 6 concludes and provides suggestions for further research.

2. Automated Intruder Detection Algorithm

2.1 Minimum Volume Sets

We expect that the set of *normal* (usual) images will constitute a high-density region of the space spanned by the set of all images. With each image constituting a multidimensional data point, the densest regions of this multidimensional space should contain the vast majority of the arriving points. Estimating Minimum Volume Sets (MVSs) is a common approach for determining high-density regions in multidimensional spaces.

Assuming that the arriving data points are drawn from a generic and unknown underlying probability distribution P, minimum volume set G_β containing mass at least $\beta \in (0, 1)$, with respect to reference measure γ, is defined as:

$$G_\beta = \arg\min\{\gamma(G) : P(G) \geq \beta\} \quad (1)$$

where G is a measurable set [6]. These sets are known in the MVS literature as density contour clusters. Estimation of MVSs satisfying (1) allows the identification of high-density regions where the mass of the underlying probability distribution is most concentrated. Points lying outside these regions may then be declared anomalous.

2.2 The One-Class Neighbor Machine

The One-Class Neighbor Machine (OCNM) algorithm proposed by Muñoz and Moguerza in [3] provides an elegant means of estimating minimum volume sets. The OCNM algorithm is a block-based procedure that provides a binary decision function indicating whether any point $\mathbf{x}_t$ is a member of a certain density contour cluster or not. The

algorithm requires the choice of a *sparsity measure*, which relaxes the density estimation problem by replacing the task of estimating the density function at each data point with a simpler measure that asymptotically preserves the order induced by the density function. Example choices for the sparsity measure include the *k*th nearest neighbor Euclidean distance and the average of the first *k* nearest-neighbor Euclidean distances.

We have implemented the OCNM algorithm here using the *k*th nearest-neighbor distance as the sparsity measure. The OCNM algorithm proceeds by sorting the values of the sparsity measure for the set of all points. It subsequently identifies those points that lie inside the MVS as those having the smallest sparsity measure, up to a pre-specified fraction μ, of the total number of points in the set.

We apply OCNM here in a sliding window manner, with the window advancing by one when a new image (when the algorithm is run locally at each node in a distributed monitoring architecture) or string of images (when the algorithm is run at the central repository in a centralized monitoring architecture) $\mathbf{x}_t$, arrives in the next timestep. The binary decision value of the algorithm output regarding the last point in the window (i.e. the most recently arrived image or string of images, $\mathbf{x}_t$) is used to flag an anomaly in real-time. Varying the pre-specified fraction μ of outliers to be isolated yields the Receiver Operating Characteristic (ROC) curves presented in Sections 4 and 5. We are then able to compare the performance of OCNM with representative algorithms from two families of algorithms commonly used for novelty detection in image sequences. The results are discussed in Sections 4 and 5.

If the *k*th nearest-neighbor distance is used as the sparsity measure and OCNM is run using a sliding window of size W, the algorithm must first evaluate the sparsity measure value of $\mathbf{x}_t$ in the window of points. This involves calculating the distance from the given point $\mathbf{x}_t$ to every other point in the window. If each point is F-dimensional, the computational complexity of this step is $\mathbf{O}(W^2F)$. The next task performed is the sorting of these W sparsity values, and most sorting functions involve an average computational complexity of $\mathbf{O}(W\log W)$ [7]. The overall computational complexity of OCNM on a window of W, F-dimensional points is thus $\mathbf{O}(W^2F)$.

3. Related Work

3.1 Principal Component Analysis

The technique of Principal Component Analysis (PCA) may be used to separate the space occupied by set of input vectors into two disjoint subspaces, corresponding to normal and anomalous behavior. An anomaly may then be flagged in the timesteps where the magnitude of the projection onto the anomalous subspace, θ_t, exceeds a threshold [8, 9, 10].

Various approaches have been proposed in literature where moving objects are detected in video sequences directly using PCA [11], and the applicability of PCA to anomaly detection in image sequences have also been suggested [12]. Wang et al. have proposed a method which uses incremental two-dimensional PCA (2DPCA) to characterize objects, with maximum-likelihood estimation used for tracking [13].

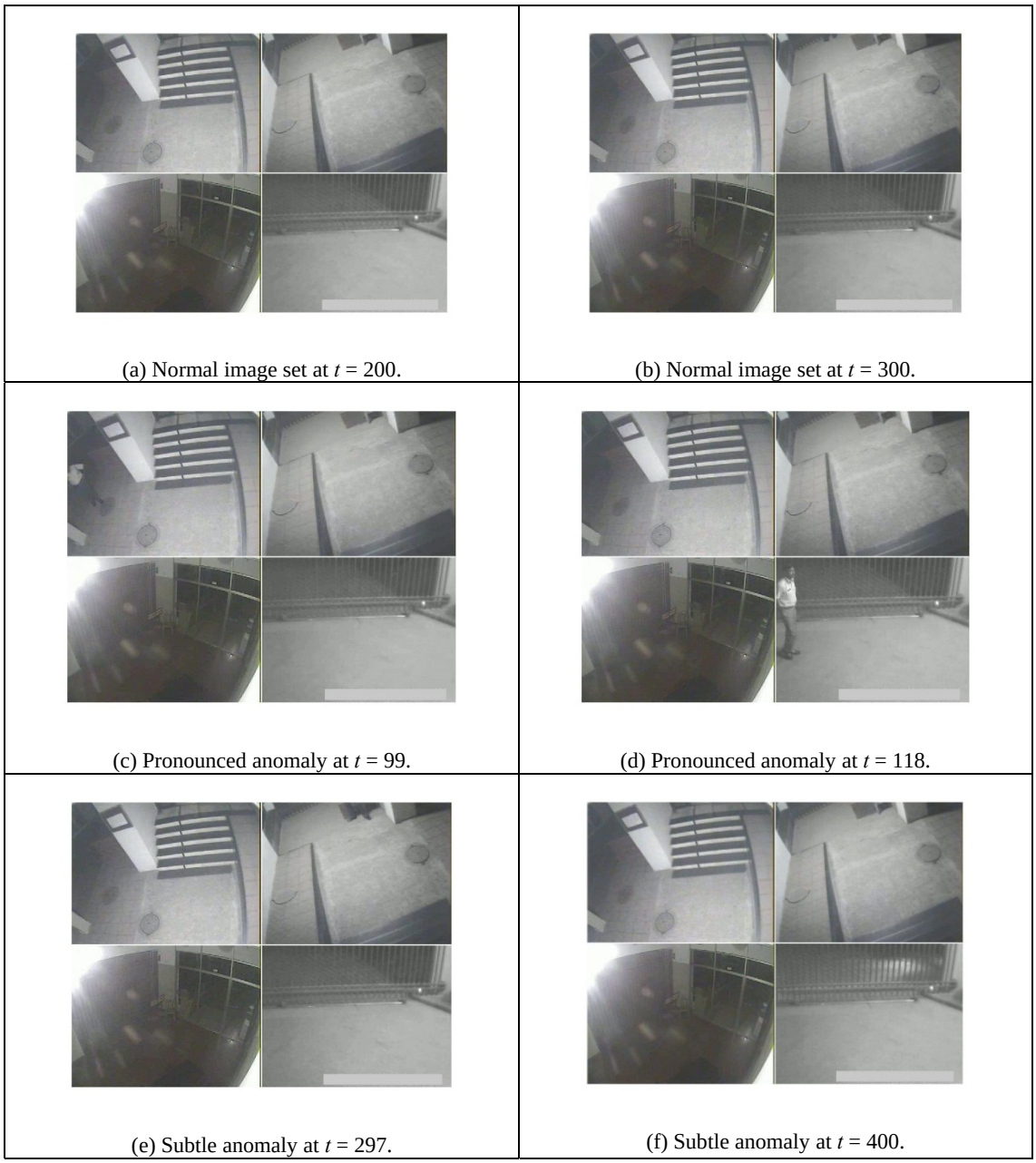
We apply PCA here in the following manner. We first verify using a *scree* plot [10] that the space is indeed over determined, and that the PCA subspace method of anomaly detection may be applied to this particular image set [12]. The number of components to be allocated to the normal and anomalous are then determined based on the *knee* in the scree plot [10, 12]. We then decide on a window size, and evaluate the magnitude of the projection of the data points onto the anomalous subspace. A binary decision regarding the last point in the window is taken by comparing the magnitude of the projection for this point, with a threshold. The window is then advanced in the next timestep as the next data point arrives, and the process is repeated. Varying the threshold yields the Receiver Operating Characteristic (ROC) curve presented in Sections 4 and 5.

Using a window of size W and assuming that the points are F-dimensional, assigning R principal components to normal subspaces provides PCA with an overall computational complexity of $\mathbf{O}(LF^2 + F^3 + RLF)$ [14], which may be simplified into $\mathbf{O}(F^3)$ where $F >> W$.

3.2 Normalized Compression Distance –based Similarity Metric

Au et al. have presented an algorithm in [15] where a set of novel images are stored, and arriving images are compared to this set. A scene is considered anomalous when the maximum similarity between the given scene and all previously viewed scenes is below a given threshold. Similarity is measured using the Normalized Compression Distance (NCD) measure [16].

The NCD measure has been shown to be a versatile and broadly applicable tool for pattern analysis, and problem formulations based on it can be very general, parameter-free, robust to noise, and portable across applications and data formats [17]. Cohen et al. have proposed an information-theoretic algorithm based on NCD to track meaningful changes in image sequences [18]. Yahalom has



(a) Normal image set at t = 200. (b) Normal image set at t = 300.

(c) Pronounced anomaly at t = 99. (d) Pronounced anomaly at t = 118.

(e) Subtle anomaly at t = 297. (f) Subtle anomaly at t = 400.

Figure 1. Set of images obtained from four cameras in the BRAC University CCTV surveillance system, corresponding to six different timesteps. Usual images are observed in two of the timesteps, two timesteps show situations where human forms are easily visible, and two show subtle cases where the foot of a person is available in one and some alien light beams are observed in another. Actual time stamps on the images have been concealed because of privacy.

developed an algorithm for web server Intrusion Detection Systems (IDS), which does not rely on signatures of past attacks, using an NCD-based metric [19].

4. Experiments: Outdoors

In this section, we present results obtained using real footage from a surveillance system that is deployed in an outdoor environment and uses centralized monitoring architecture.

4.1 Data

We collected real footage from a set of four cameras from the centralized CCTV network in place at BRAC University. The raw data is comprised of a concatenated video sequence in the AVI format. From the videos, we extracted (concatenated) still images in the JPEG format at two-second intervals. The total data set consisted of 500 timesteps, of which 62 were identified as potential anomalies after performing an exhaustive, manual inspection of the data set.

Figure 1 shows pictures corresponding to six example timesteps. Within Fig. 1, subfigures (a) and (b) show regular (normal) scenarios; (c) and (d) show obvious cases of human forms appearing on the scene (top-left and bottom-right cameras, respectively); (e) presents a subtle case where a small portion of a person's foot is visible (top-

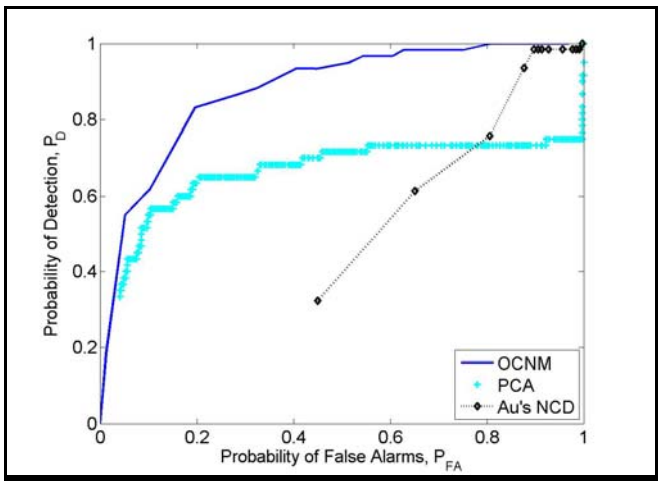


Figure 2. ROC curves showing performances of proposed OCNM versus existing PCA and Au's NCD-based algorithms. OCNM is seen to substantially outperform PCA and NCD. Experiments conducted on outdoor images from BRAC University using centralized monitoring architecture.

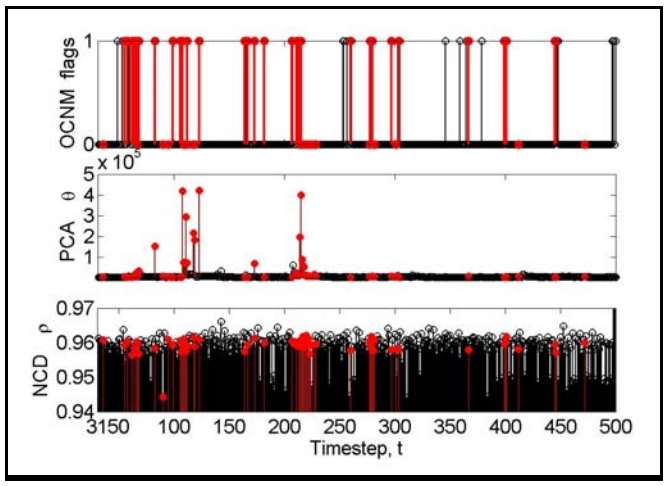


Figure 3. Progression in the anomaly detection statistics for each algorithm for an example setting of the relevant detection thresholds. Top panel: Timesteps flagged by OCNM. Middle panel: Magnitude of projection onto the residual subspace, θ_t, for PCA. Bottom panel: $1–\rho_t$, where ρ_t is Au's NCD-based similarity metric. The true anomalies are indicated as red stems with filled circles. Experiments conducted on outdoor images from BRAC University using centralized monitoring architecture.

right camera); (f) presents another subtle case where alien lights appear (bottom-right camera). To be conservative, we also identify scenarios such as (e) and (f) as potential anomalies that the operator may wish to be alerted to. The actual timestamps have been removed from the images for the sake of privacy.

4.2 Feature Extraction using Wavelet Decomposition

After extracting JPEG images at two-second intervals from the AVI video, we performed standard two-dimensional Haar wavelet decompositions to obtain a 120 × 160 × 3 tensor representation for each image. Working in the frequency domain is preferable to working in the space domain in order to account for differences between specific pixels in different images arising as a result of minor camera movements, and also to consider and compare each image as a whole. Wavelets provide a convenient technique for representing image details in the frequency domain. The wavelet decomposition represents an image in a manner that reflects variation in neighboring pixel intensities, and also performs image compression. Because this representation relates neighboring pixel intensities, it is also suitable to be fed into algorithms which look to find patterns between higher order statistics of the pixels.

We finally performed 10% bilinear interpolation to rescale and reduce the size of each dimension. The output of the four cameras was then concatenated to obtain one 120 × 160 × 3 × 4 = 2304-dimensional row vector of input data corresponding to each timestep.

4.3 Results

Figure 2 compares the performances of OCNM with PCA and Au's NCD-based algorithms through ROC curves, demonstrating the tradeoff between the Probability of False Alarms (P_{FA}) and the Probability of Detection (P_D). The curves were obtained by varying the anomaly detection thresholds for each algorithm. A window size of 30 was used. For OCNM, the nearest-neighbor parameter k was set to two. For PCA, four principal components were assigned to the normal subspace, while Au's NCD-based algorithm was run using the author's recommended settings from [20]; these yielded the best results for PCA and NCD. It is evident from Fig. 2 that the performance of OCNM is substantially superior to the performances of PCA and NCD. Moreover, OCNM is easily able to achieve near-perfect detection rates. The low performance of the NCD algorithm may be attributed to the fact that this algorithm requires a significantly longer training period, and needs to maintain a significantly larger database of images to compare new arrivals against [15, 20].

Figure 3 (top panel) shows the particular timesteps that OCNM flags as anomalous, using the representative value of $\mu = 0.90$ set to identify the 10% outliers. The locations of the "true" anomalies, as we manually identified, are indicated as red stems with filled circles. Comparison with PCA (middle panel) and NCD (bottom panel) indicates that OCNM does the best job of isolating the identified anomalies, in agreement with the ROC curves from Fig. 2.

5. Experiments: Indoors

In this section, we present results from a controlled experiment on a network of cameras setup in an indoor environment. Each camera processes the images and runs the detection algorithms locally, thus employing a distributed monitoring architecture.

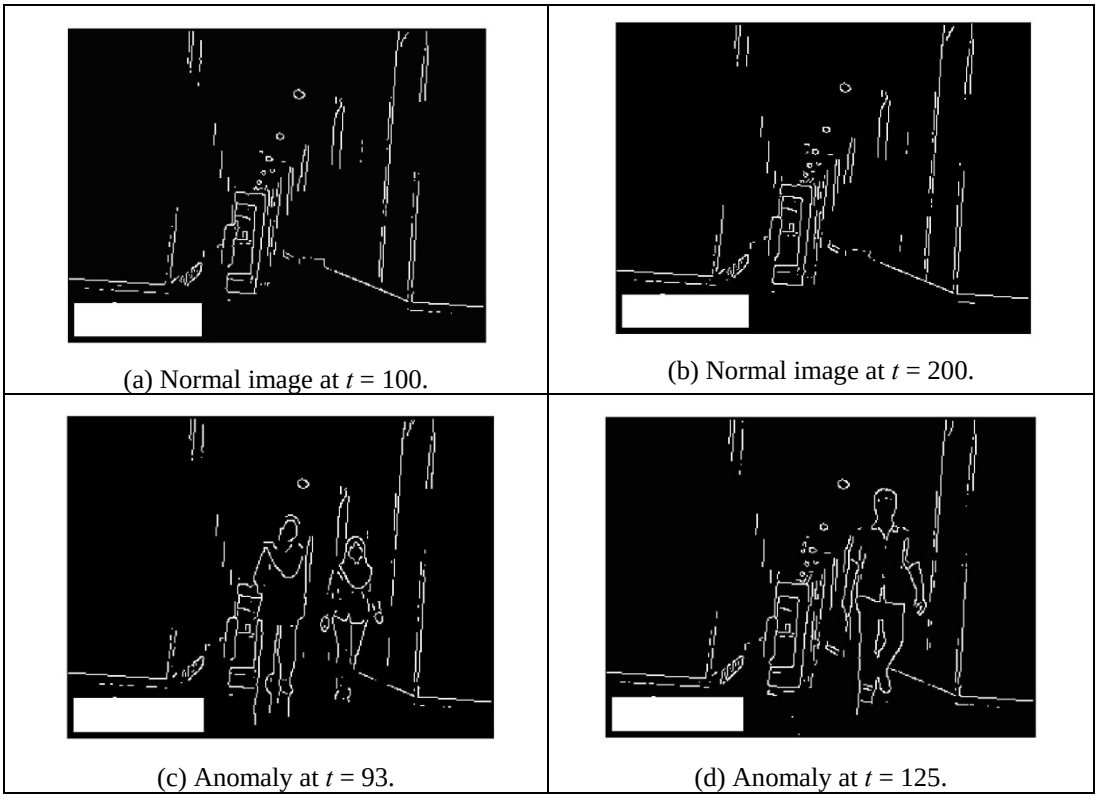
5.1 Data

We setup a network of four Logitech™ webcams in a junction of hallways at the Department of Computer Science at International Islamic University Malaysia (IIUM). Each webcam was connected to a Dell laptop computer which

(a) Normal image at t = 100. (b) Normal image at t = 200.

(c) Anomaly at t = 93. (d) Anomaly at t = 125.

Figure 4. Images captured during four different timesteps by Camera 1 in an indoor network of webcams set up at IIUM. Two timesteps show usual images and two show situations where humans appear. Actual time stamps on the images have again been concealed because of privacy.



(a) Normal image at t = 100. (b) Normal image at t = 200.

(c) Anomaly at t = 93. (d) Anomaly at t = 125.

Figure 5. Canny edge images corresponding to the raw images from Camera 1 at IIUM that were presented in Fig. 4. Extra edges are visible in (c) and (d), corresponding to the extra physical forms present here.

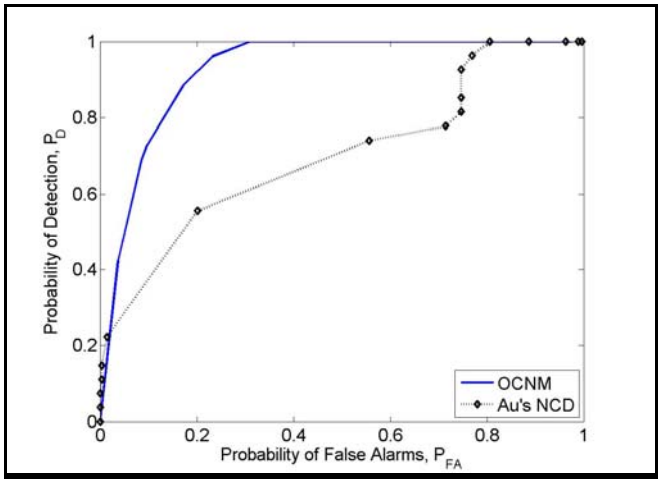


Figure 6. ROC curves showing performances of proposed OCNM versus Au's NCD-based algorithm. OCNM is seen to substantially outperform NCD. Images from Camera1 in the indoor network with distributed monitoring architecture setup at IIUM.

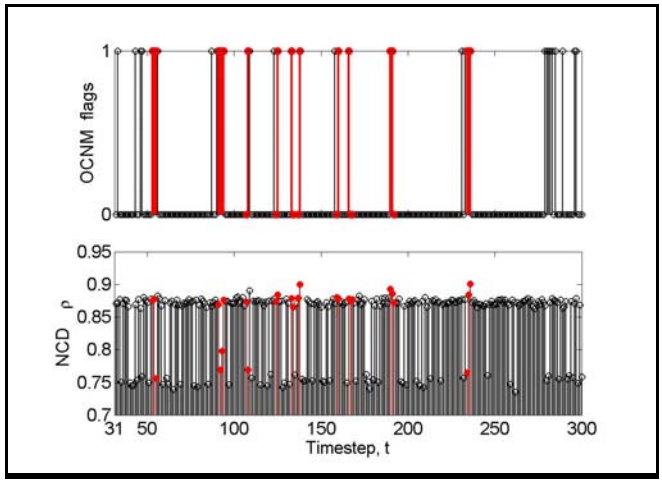


Figure 7. Progression in the anomaly detection statistics for each algorithm for an example setting of the relevant detection thresholds. Top panel: Timesteps flagged by OCNM. Bottom panel: $1-\rho_t$, where ρ_t is Au's NCD-based similarity metric. The true anomalies are indicated as red stems with filled circles. Images from Camera1 in the indoor network with distributed monitoring architecture setup at IIUM.

was programmed to take still snaps at every 15-second interval and individually run the detection algorithms. This setup simulated a distributed monitoring architecture, with higher resolution cameras, in an indoor environment. The raw data thus comprised of still images in JPEG format at 15-second intervals. The total data set consisted of 300 timesteps, of which 27 were identified as potential anomalies after again performing an exhaustive, manual inspection of the data set.

Figure 4 shows pictures from Camera 1 corresponding to four example timesteps. Within Fig. 4, subfigures (a) and (b) show regular (normal) scenarios, and (c) and (d) show instances of human forms appearing on the scene. The actual timestamps have again been removed for the sake of privacy.

5.2 Pre-processing using the Canny Edge Detector

A high-resolution camera captures a lot of details which may be unnecessary given our objective of detecting physical intruders in the camera's field of vision. An elegant way of eliminating such noise is provided by the Canny edge detector [5]. The Canny edge detector filters out information such as illumination and hue from the original image, while preserving the important structural properties [15]. The result is an "edge" image where the step edges are enhanced. Figure 5 shows the Canny edge images corresponding to the example raw images from Fig. 4. The edge images instantly draw our attention to the essential difference between subfigures (c) & (d) from subfigures (a) & (b), in the additional physical shapes present in subfigures (c) & (d).

5.3 Feature Extraction using Wavelet Decomposition

After obtaining the Canny edge images (also in the JPEG format) at 15-second intervals, we again performed standard two-dimensional Haar wavelet decompositions to obtain a 120 × 160 vector representation for each image at each camera, corresponding to each timestep. Transformation into the frequency domain is still necessary to account for differences between specific pixels in different images arising as a result of minor camera movements, and to consider and compare each image as a whole.

We finally performed 25% bilinear interpolation to rescale and reduce the size of each dimension. This yielded one 30 × 40 = 1200-dimensional row vector of input data at each camera corresponding to each timestep.

5.4 Results

Figure 6 compares the performances of OCNM with Au's NCD-based algorithm through ROC curves, demonstrating the tradeoff between the Probability of False Alarms (P_{FA}) and the Probability of Detection (P_D). The curves were obtained by varying the anomaly detection thresholds for each algorithm. Again, a window size of 30 was used, the OCNM nearest-neighbor parameter k was set to two, and Au's NCD-based algorithm was run using the author's recommended settings [20]. It is evident from Fig. 6 that the performance of OCNM is again substantially superior to that of NCD. Moreover, OCNM is again easily able to achieve near-perfect detection rates. It may also be observed that both OCNM and NCD show relatively better performances here compared to the outdoor image sequence collected from the BRAC University CCTV network (Section 4). This is because the raw images here are of higher resolution. PCA yielded unacceptably low detection performance in this indoor experiment where we use wavelet transforms of the Canny edge images (instead of wavelet transforms of the raw JPEG images) as the image representation. Hence, we do not present results using PCA here.

Figure 7 (top panel) shows the timesteps that OCNM signals as anomalous, again for the representative setting of $\mu = 0.90$ identifying the 10% outliers. The locations of the "true" anomalies, as we manually identified, are indicated as red stems with filled circles. Comparison with NCD (bottom panel) indicates that OCNM a better job of isolating the identified anomalies, in agreement with the ROC curves from Fig. 6.

6. Conclusions and Future Work

In this paper, we have presented a novel approach to performing real-time intruder detection in a surveillance system using inexpensive components. We have proposed the One-Class Neighbor Machine algorithm that is based on the theoretical concept of Minimum Volume Sets. We have demonstrated high detection rates and performance superior to representative existing algorithms through applications to image sequences from two different scenarios: a real, run-of-the-mill, already-deployed Closed-Circuit Television outdoor surveillance system employing a centralised monitoring architecture, and a camera network specifically constructed indoors using a distributed monitoring architecture.

Our future work will focus on integrating face detection algorithms to learn the characteristics of the regular visitors to the applicable premises [21]. In addition, we wish to investigate other real-time, adaptive anomaly detection algorithms such as Kernel Estimation-based Anomaly Detection (KEAD) [14], and explore different Principal Component Analysis (PCA) variants [22].

Acknowledgement

This project was funded in part by the Research Management Centre, International Islamic University Malaysia under grant number EDW B11-167-0645.

References

[1] M. Valera and S. Velastin, "A review of the state-of-the-art in distributed surveillance systems," in *Intelligent distributed video surveillance systems*, S. Velastin and P. Remagnino, Eds. London, UK: Institution of Electrical Engineers, 2008, pp. 1–30.

[2] S. Velastin & P. Remagnino, *Intelligent distributed video surveillance systems*. London, UK: Institution of Electrical Engineers, 2008.

[3] A. Muñoz and J. Moguerza, "Estimation of high-density regions using one-class neighbor machines," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 28, no. 3, pp. 476–480, Mar. 2006.

[4] T. Ahmed, X. Wei, S. Ahmed, and A.-S. Pathan, "Intruder detection in camera networks using the one-class neighbor machine," in *Proc. American Telecommunications Systems Management Association (ATSMA) Networking and Electronic Commerce Research Conf.* (*NAEC*), Riva del Garda, Italy, Oct. 2011.

[5] J. Canny, "Computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.

[6] J. Einmal and D. Mason, "Generalized quantile processes," *Annals of Statistics*, vol. 20, no. 2, pp. 1062–1078, Jun. 1992.

[7] T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Cambridge, MA, USA: MIT Press, Sep. 2001.

[8] T. Ahmed, M. Coates, and A. Lakhina, "Multivariate online anomaly detection using kernel recursive least squares," in *Proc. IEEE Int. Conf. on Computer Communications (INFOCOM)*, Anchorage, AK, USA, May 2007.

[9] T. Ahmed, B. Oreshkin, and M. Coates, "Machine learning approaches to network anomaly detection," in *Proc. ACM/USENIX Workshop on Tackling Computer Systems Problems with Machine Learning Techniques (SysML)*, Boston, MA, USA, Apr. 2007.

[10] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. Kolaczyk, and N. Taft, "Structural analysis of network traffic flows," in *Proc. ACM SIGMETRICS*, New York, NY, USA, Jun. 2004.

[11] N. Verbeke and N. Vincent, "A PCA-based technique to detect moving objects," in *Image Analysis*, B. Ersbøll and K. Pedersen, Eds. Berlin/Heidelberg, Germany: Springer, Jul. 2007, vol. 3633/2005, pp. 641–650.

[12] T. Ahmed, S. Ahmed, S. Ahmed, and M. Motiwala, "Real-time intruder detection in surveillance systems using adaptive kernel methods," in *Proc. IEEE Int. Conf. on Communications (ICC)*, Cape Town, South Africa, May 2010.

[13] T. Wang, I. Gu, and P. Shi, "Object tracking using incremental 2DPCA learning and ML estimation," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, Honolulu, HI, USA, Apr 2007.

[14] T. Ahmed, "Online anomaly detection using KDE," in *Proc. IEEE Global Communications Conf.* (*GLOBECOM*), Honolulu, HI, USA, Nov. 2009.

[15] C. Au, S. Skaff, and J. Clark, "Anomaly detection for video surveillance applications," in *Proc. IEEE Int. Conf. Pattern Recognition (ICPR)*, Hong Kong, China, May 2006.

[16] M. Li, X. Chen, X. Li, B. Ma, and P. Vitanyi, "The similarity metric," *IEEE Trans. Information Theory*, vol. 50, no. 12, pp. 3250–3264, Dec 2004.

[17] E. Keogh, S. Lonardi, and C. Ratanamahatana, "Towards parameter-free data mining," in *Proc. ACM SIGKDD*, Seattle, WA, USA, Aug. 2004.

[18] A. Cohen, C. Bjornsson, S. Temple, G. Banker, and B. Roysam, "Automatic summarization of changes in biological image sequences using algorithmic information theory," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 31, no. 8, pp. 1386–1403, Aug. 2009.

[19] S. Yahalom, "URI anomaly detection using similarity metrics," Master's thesis, Tel-Aviv University, Tel Aviv, Israel, May 2008.

[20] C. Au, "Compression-based anomaly detection for video surveillance applications," Master's thesis, McGill University, Montreal, QC, Canada, Feb. 2006.

[21] M. Faruqe and M. Hasan, "Face recognition using PCA and SVM," in *Proc. IEEE Int. Conf. on Anti-counterfeiting, Security, and Identification in Communication* (*ASID*), Hong Kong, China, Aug. 2009.

[22] X. Wei, T. Ahmed, M. Chen, and A.-S. Pathan, "PeerMate: A malicious peer detection algorithm for P2P Systems based on MSPCA," in *Proc. IEEE Int. Conf. on Computing, Networking and Communications* (*ICNC*), Lahaina, HI, USA, Jan. 2012.