A Pixel-level, Intensity-based Nonlinear Autoregressive Classifier (NARX) with Chromatic Exogenous Input for Efficient Image Background Subtraction

Syed A. Yusuf, David J. Brown, Alan Mackinnon, and Richard Papanicolaou

Abstract— **Background subtraction is a well-known technique in computer vision to extract foreground objects from background reference frames. In real-time video processing applications such as surveillance, behavioral profiling and intelligent transport systems, the domain presents a number of challenges. Video frames used to train such models contain a range of dynamic background activities such as waving trees, moving cloud cover or abrupt intensity variations that make the foreground detection a challenging task.**

Dynamic neural networks are known for their capability to predict time-series-based nonlinear models via previous feature data. The proposed scenario models each pixel's intensity/color-alternating behavior based on its previous activity patterns. Any significant or unusual variation in the underlying intensity or color value therefore is modeled as a foreground activity. Based on this concept, this paper presents a non-linear autoregressive neural (BG-NARX) classifier with the pixels' chromatic values as the exogenous vectors to improve background detection accuracy.

The proposed model was evaluated against three benchmarking video datasets and reported promising detection accuracies ranging from 67-94% for pedestrians and vehicles against highly variable backgrounds with low false positives and negatives.

I. Introduction

With a global increase in the requirements of automated security and surveillance applications, the role of behavioral biometrics is getting extremely important. In real-time video processing the capability of computer vision techniques to detect stationary objects via static cameras is regarded as the first step in an efficient recognition of foreground objects. The area has a wide range of applications in video analytics inclusive of automated CCTV security and surveillance, human behavioral profiling and traffic monitoring in intelligent transport systems. Majority of such cases suffer from a substantial level of "background activity" in the form of waving trees, water flow, moving cloud-cover and/or abrupt intensity variations. Due to the presence of these "unwanted anomalies", classification of genuine foreground objects such as vehicles or pedestrians remains a challenging task. Lately, with the ever-increasing role of gesture and face profiling applications in handheld and smart devices, the domain has attracted an increased focus with research increasingly converging on pixel-level and region modeling of video frames.

Background subtraction is a specialist image segmentation methodology aimed to eliminate dynamic and static background information where the underlying model is being trained on image sequences taken from a set of previous time instances or pixel profiles. Statistical analysis of pixel-level intensity and color-distortion modeling is a well-known technique in background subtraction. The technique utilizes RGB value distribution of pixels as time-based Gaussian mixtures or as minimum/maximum intensity differences between subsequent frames [1, 2]. These methods often fail in the presence of a highly dynamic multimodal background. Multimodal pixel behaviors generally arise when different objects occasionally block the same pixel location at different instances of time such as a waving flag against a building. The problem has widely been reported in literature though majority of techniques suffer from false segmentations due to global illumination changes.

Gaussian mixture models were initially used by Stauffer & Grimson [3] to treat each pixel in an image matrix as a mixture model trained over time. Statistical Bayesian estimation was later used by Lee *et al*. [4] and Harville [5] and utilized time-adaptive mixture modeling to compensate for new changes appearing within the scene with the older-ones being discarded over-time. Though the methodology efficiently modeled and discarded repetitive scenes such as the cloud-cover or moving dials of a clock, it suffered substantially from a learning-rate adjustment issue where the model was incapable of incorporating sudden intensity changes. The issue can be resolved in majority of cases, however, a higher learning rate resulted in the slow-moving-backgrounds being absorbed in the background. Moreover, GMM-based background subtraction technique is also

Manuscript received March 21, 2013. This work was supported by the Knowledge Transfer Partnerships/Technology Strategy Board, UK. The project is a joint venture between STS Defence Ltd and the University of Portsmouth.

Yusuf A. S. is with the STS Defence Ltd, Gosport, PO12 1AF UK (phone: 02392 584222; fax: 02392 529598; e-mail: adnan.yusuf@sts-defence.com).

Mackinnon. A. is with the STS Defence Ltd, Gosport, PO12 1AF UK (phone: 02392 584222; fax: 02392 529598; e-mail: amackinnon@sts-defence.com).

Brown D. J. is with the Institute of Industrial Research, University of Portsmouth, Portsmouth, PO1 2EG UK (phone: 02392 844446; e-mail: david.j.brown@port.ac.uk).

Papanicolaou R. is the Managing Director at STS Defence Ltd, Gosport, PO12 1AF UK (phone: 02392 584222; fax: 02392 529598; e-mail: richardp@sts-defence.com).

known to erroneously incorporate shadows as foreground objects which is a significant issue in behavioral monitoring domain. In order to eliminate these shortcomings, Kim *et al*. [6] proposed a pixel-level codebook-based technique utilizing the maximum negative runtime length (MNRL) parameter of time-series-based pixels to identify background pixels from foreground pixels. The technique did manage a promising accuracy against intensity varying scenes. Adding further to this work, Ilyas *et al*. [7] introduced frequency-based attributes in the codebook in order to add, delete or match codewords from the codebook. Though these techniques present a substantial advantage over conventional GMM-based background subtraction, there still exists an additional memory overhead which is directly proportional to the number of discrimination variables being kept in the memory. Recently, Yusuf *et al*. [8] presented an intensity Euclidean-distance and kurtosis-based methodology to reduce the memory overhead involved to incorporate objects re-appearing in the scene at longer time intervals. Video sequence assessment at its core is a time-series analysis where gradual or recurrent changes in a pixel's chromaticity or intensity values overtime is to be labeled as a normal background activity. Any abrupt change, however, such as in the case of an external object, must be labeled as a foreground process.

Artificial neural networks (ANNs) are well-known in their capability to model recurrent noise signals (such as tree leaves or water fountains). The technique has generally been applied in nonlinearly-dynamic real-world systems to predict financial parameters [9] and weather forecasting [10]. The technique has frequently been applied to pixel- and image-level time-series modeling in computer vision. Generalized regression-based neural networks were used by Luv *et al*. [11] by combining a set of dynamic signals by processing the feature space through discrete cosine transforms. Dynamic neural networks were similarly used to detect spatio-temporal action detection to minimize false identification and improve foreground detection via recursive Bayesian learning [12, 13]. The capability of ANN to incorporate historic changes into account to predict an unknown variable at time t based on the previous n time instances makes it an ideal platform to model nonlinear behavior. Dynamic multi-modal pixel behavior is a similar nonlinear activity that can be used to predict the future behavior of that pixel in terms of its intensity or chromatic information change.

Based on this capability of dynamic time series neural networks, the paper proposes results from the application of a nonlinear autoregressive classifier with external input. The paper is organized as follows: Section II presents two core areas of development in this research: namely the video image feature extraction methodology and the subsequent neural training technique for background modeling. Section III presents the outcomes and analysis of the results against a set of standard video benchmarking datasets. Section V finally concludes with the planned extensions of the proposed methodology.

II. METHODOLOGY

A. A vector-angle/skew-based background subtraction methodology

Vector Angle (VA) and Euclidean Distance (ED)-based models are commonly used to create color invariant segmentation or edge-detection. The model was initially proposed by Dony and Wesolkowski [14] to implement an intensity/color-invariant segmentation algorithm. As shown in Fig. 1, based on the ability of this model to discriminate between variable colors as well as the intensity changes, the degree of chromaticity and intensity variation for each pixel can be used to model pixel-level variation over a time-series.

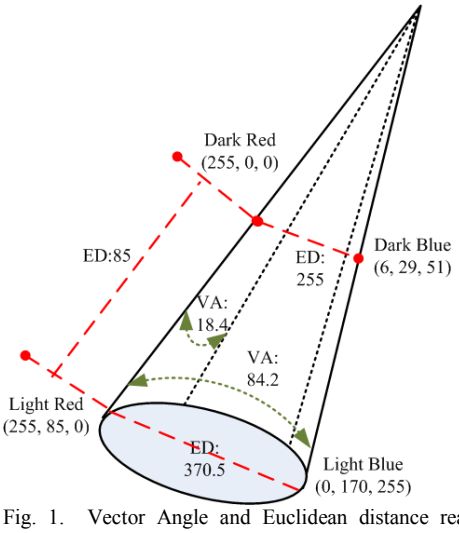

Fig. 1. Vector Angle and Euclidean distance realization between two different color- and intensity-spaces

For the pixel $p(i,j)$ where i and j are the row and column indices, the ED and VA mapping for a color and intensity spectrum shown in Fig. 1 is given in (1) and (2):

$$ED_t(p_{i,j}) = \|\overline{RGB}_t - \underline{RGB}_t\| \quad (1)$$

In (1), $\|\circ\|$ is the L_2vector norm, $RGB = [\![r\ g\ b]\!]^T$ and $\overline{RGB}$ and $\underline{RGB}$ are maximum and minimum values noted over a time duration $0 \leq t \leq T$ where T is the training-time duration set to 20 seconds.

Moreover, colors that only differ in intensity tend to stay at collinear locations on the map shown in Fig. 1 whereas those differing in hue and/or saturation are separated by the angle θ calculated as follows in (2):

$$cos\theta = RGB_1^T - RGB_2 / \|RGB_1\| \|RGB_2\| \quad (2)$$

Fig. 2 (a) represents a slow and gradual intensity decline in the global light intensity within the scene. The dataset is taken from the Microsoft Wallflower benchmarking dataset with (b) and (c) representing two different frames from a set of 500 frames [15, 16]. The ED and VA maps shown in (d) and (e) represent a direct relationship between abrupt

intensity and color changes due to the pixel (shown by '*') value being altered due to a person moving in (shown in c). However, the point worth noting in VA map shown in (d) is its failure to detect an angular change at a very low intensity and was only able to detect the change at frame 51. Nonetheless, the outcome demonstrates the suitability of ED and VA-based measure to record abrupt as well as slow scene changes. The ED/VA values thus obtained were further evaluated in an outdoor water fountain scene from the Wallflower database against a highly dynamic background to prove the effectiveness of the statistical skew measure given in (2).

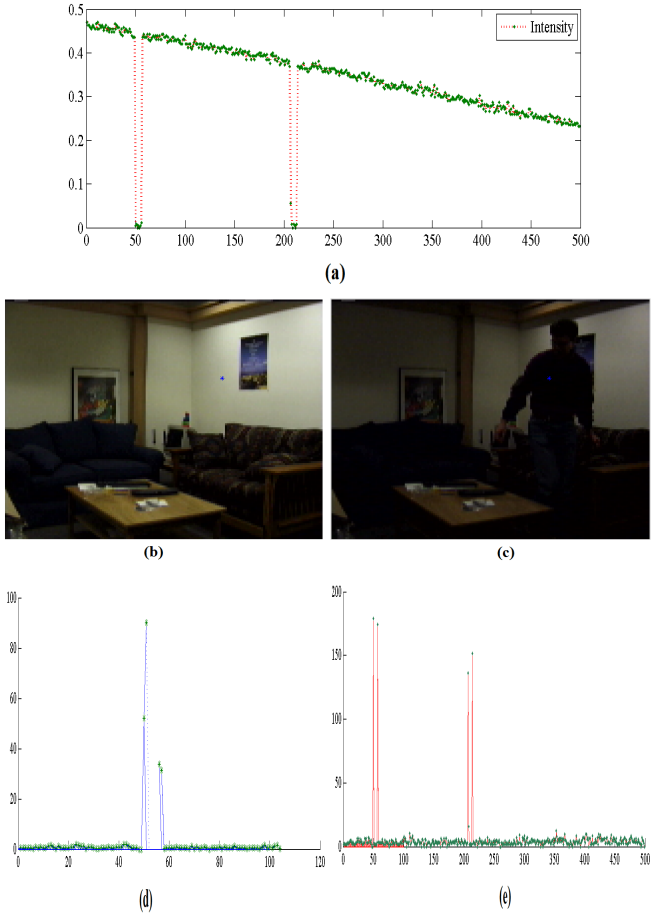

Fig. 2. (a) Demonstrates an overall intensity drop due to a simulated day-time change induced for the Wall Flower dataset [15, 16] with (b) and (c) representing variable lighting conditions and '*' marking the pixel location tested over a set of 500 frames. (d) and (e) represent the intensity and chromaticity change behavior based on ED and VA application respectively

B. Foreground codebook replacement based on a dynamic neural network model with exogenous vector input

Traditional and recent codebook methodologies predominantly rely on updating codebooks for each pixel in an image intensity/chromaticity matrix where for a training sequence $\rho(x) = \{x_1, x_2, \dots, x_N\}$, a codebook $\delta = \{\delta_1, \delta_2, \dots, \delta_M\}$ is kept where the size $\mathbf{M}$ depends upon the underlying sample variance for the pixel ρ. The codebook entries are updated based upon a multi-tuple auxiliary variable where a certain pixel intensity/chromaticity variation is added in the background codebook only when one or more of the following parametric bounds are violated:

- $\rho_{t,j}(\hat{\vartheta}, \check{\vartheta})$: The minimum and maximum brightness values for the pixel at t^{th} row and j^{th} column.
- $\rho_{t,j}(\phi)$: The frequency with which a certain pixel variation occurs
- $\rho_{t,j}(\eta)$: The maximum negative runtime length for $\rho_{t,j}$ for which its intensity/chromaticity value has not changed beyond a threshold T
- $\rho_{t,j}(t_l, t_f)$: The first and latest codebook access times

Despite presenting good foreground classification outcomes, the larger variable space for a complex $\mathbf{m} \times \mathbf{n} \times \mathbf{L}$ codeword space further multiplied by the memory requirement for each of the abovementioned variables increases the memory requirement of the underlying algorithm.

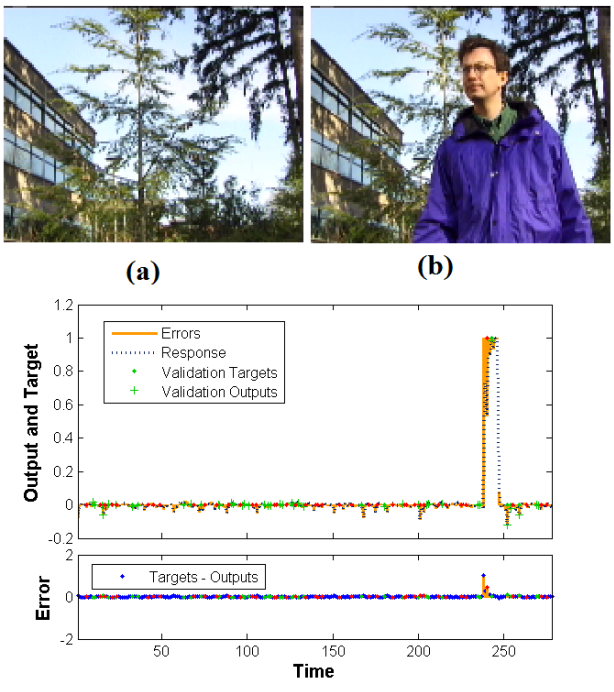

Fig. 3. (a) Frame 60 and (b) 247 of a highly dynamic background "waving trees" dataset from Wallflower benchmarking dataset for pixel location 100, 60 with (c) representing the NARX training time-series response against the proposed EA/VD-based 7-second time-delay input,

In order to reduce the memory-intensity requirement of codebook-based techniques, the proposed methodology predominantly focuses on a simple time-series network. The model is based on ED and VA-based pixel foreground/background estimation which relies on a time-delay-based identification of pixels based on previous intensity/color variation history.

In the proposed technique, a nonlinear autoregressive neural classifier is used based on intensity-based ED time-series data and color-information-based VA exogenous input. The training data was taken from the Microsoft Wallflower "waving trees" dataset which was manually labeled for foreground and background pixel classifications from frame 1 to 250. The data contained 285 RGB video frame sequences of 160×120 pixel resolution where the remaining 35 frames were used for model evaluation over unseen frame data.

Due to a very sparse foreground dataset, the system was trained over 30% training data with 35% testing and validation samples chosen randomly from the respective dataset. The network thus created was trained in an open-loop form (See Fig. 4) which provides the capability to supply the network with correct past outputs during the training process thereby further increasing the prediction process. The "waving trees" training data is shown in Fig. 5. From Fig. 5 (c) training-error-plot, it must be noted that there was a gradual error improvement as the person, starting from frame 245, alters the intensity/chromaticity values of pixel (100, 60).

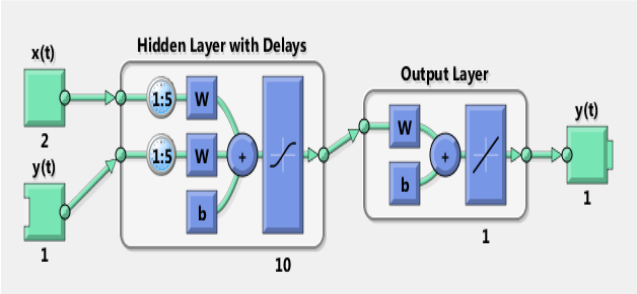

Fig. 4. The nonlinear autoregressive architecture for background modeling with $x(t)$ as a two-column feature vector containing ED/VA time-series based over the previous 5-frame data and an open-loop output feedback

TABLE I. ED/VA MODEL TRAINING OUTPUT-TARGET CORRELATION BASED ON 250 SAMPLE FRAMES FROM THE "WAVING TREES" DATASET

Model/Delay	Data Samples	MSE $\times 10^{-2}$	Regression
Training	50	0.0976	0.987
Validation	100	1.18	0.831
Testing	100	1.32	0.78

The initial "waving trees" model was trained with different time-delay values and a five-step-delay was taken as the best performing system. The training outcome was repetitively evaluated against various time-delays and terminated when validation showed six consecutive validation failures.

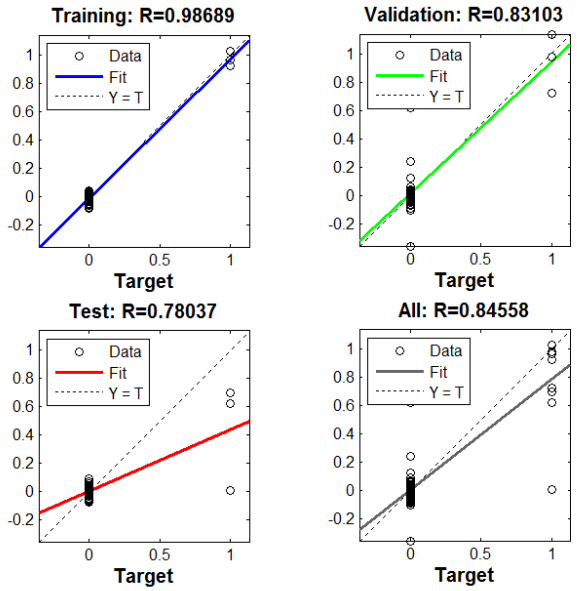

Fig. 5. Regression outcome for the "waving trees" dataset demonstrating a robust target-to-outcome relationship particularly for validation and training data

Moreover, in order to compensate for a smaller foreground dataset, the training was iteratively performed until the regression outcome of validation dataset improved beyond 0.8. The training MSE and output-target regression relationship is shown in Table I. The low validation error values demonstrate the robustness of the underlying neural classifier to incorporate tree-movement in the background. The regression plot, particularly for training and validation data also demonstrate the network's identification capability as shown in Fig. 5. The trained classifier was ultimately used to evaluate the remaining 35 video frames in addition to two alien datasets. In an ideal condition, separate training must be done in the immediate training case in order to improve the accuracy of the classifier. As the original datasets are manually labeled, for complex video scenes as that of MIT video traffic file discussed in Dataset Three in Section III, manual labeling is a cumbersome process.

III. OUTCOMES AND ANALYSIS

The trained NARX classifier was evaluated against three benchmarking datasets containing major background challenges. The datasets were evaluated against the proposed BG-NARX model with a 5-frame-delay and 10 hidden neurons with vector-angle change feature (VA) as the exogenous input (Shown in Fig. 4 as $y(t)$). The details of these datasets are given as follows:

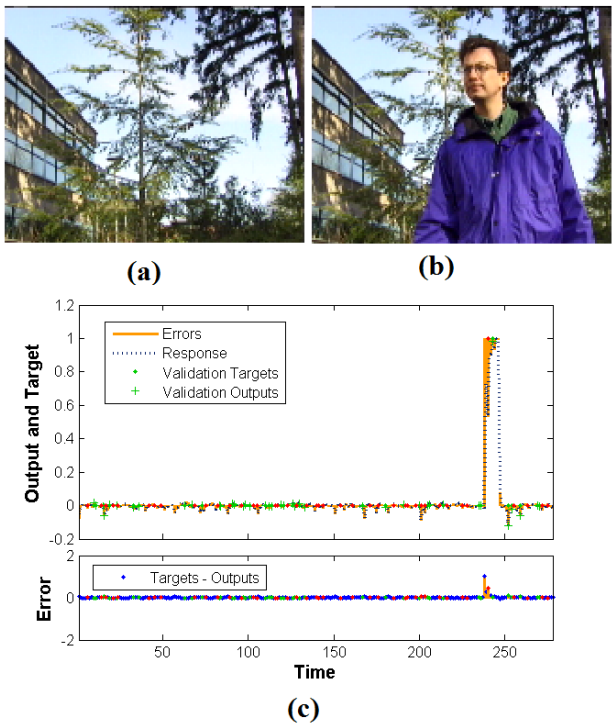

Fig. 6. (a) Frame 60 and (b) 247 of a highly dynamic background "waving trees" dataset from Wallflower benchmarking dataset for pixel location 100, 60 with (c) representing the NARX training time-series response against the proposed EA/VD-based 5-second time-delay input [Note]: It must be noted that outcome from frame 250 onwards in this figure shows the evaluation against the data unknown to the classifier

A. Dataset one (Waving trees)

The "waving trees" dataset from Wallflower consists of 286 frames of bitmap type taken at a resolution of 160×120, recorded at 96 dpi and a 24 bit depth. The earlier 250 frames of this dataset were used to train the underlying BG-NARX model whereas frames 251 to 286 were treated as unseen data. The core challenge in this dataset was a highly dynamic but monotonous motion of tree leaves. The pixels shown in Fig 7 (b) being falsely classified as background pixels indicates the inability of the NARX classifier to properly model the foreground model due to a lack of foreground data. Fig. 7 demonstrates the evaluation of BG-NARX model against the unseen "waving trees" data on frame 255. It must be noted that the frame is part of the "unseen" sample frames that were not originally used in the training sample. The frame shows a large number of false positives and negatives which can directly be attributed to an imbalanced and sparsely distributed dataset. It must be noted that only 8, manually labeled instances (from frame 243 to 250) were used to model a foreground situation compared to 242 background training frames.

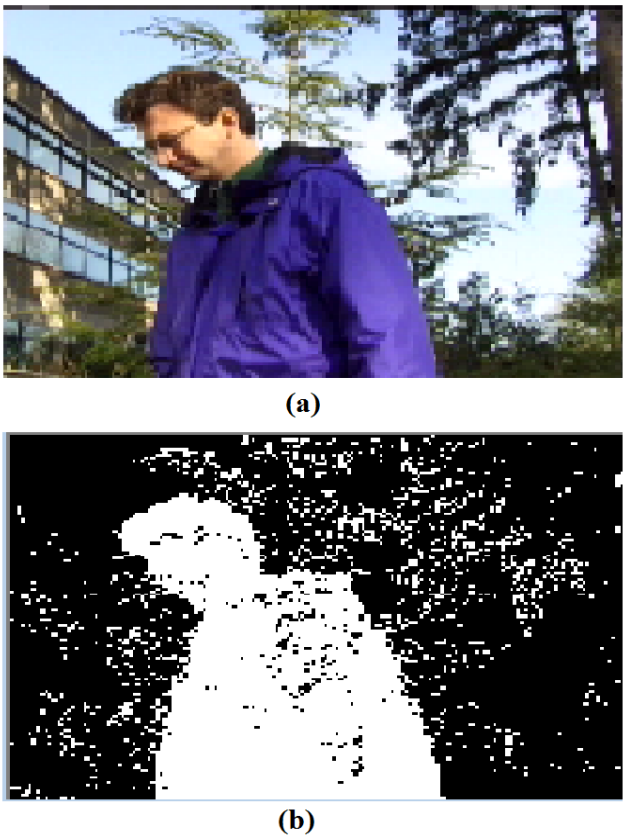

Fig. 7. Proposed BG-NARX foreground/background outcome on (a) original frame 255 and (b) proposed BG-NARX classification

B. Dataset two (Bus station at dusk)

The Wallflower "bus-station" dataset contained three main challenges namely glass reflection from dark-tinted walls, building/pedestrian shadows and rapidly changing lighting conditions due to dusk. The ED/VA mapping of this case is shown in Fig. 8. The dataset consists of 1250 frames of JPG type taken at a resolution of 360×240 recorded at 96 dpi and a 24 bit depth. In Fig. 6 it must be noted that the person on the left got absorbed into the background due to an extended period of inactivity (5+ seconds). The person on the right, however, had only recently started moving which was robustly incorporated as a foreground representation. The methodology also suffered from false positives/negatives (See Table III) generated due to shadows which can further be seen in Fig. 6 (a) as well as (c). In these two cases a false positive occurred at the 1090th frame when the shadow of the person-on-the-right passed through pixel 'o'. Moreover, a false negative was detected when the person-on-the-left stayed dormant for a few seconds thereby absorbing in the background. It must be noted that the Euclidean values for shadows are generally lower than the actual object due to light only partially blocked by shadows contrary to actual objects. Though, the outcome does not resolve the shadow problem, it does promote a promising avenue for shadow detection and modeling as a future extension to this work.

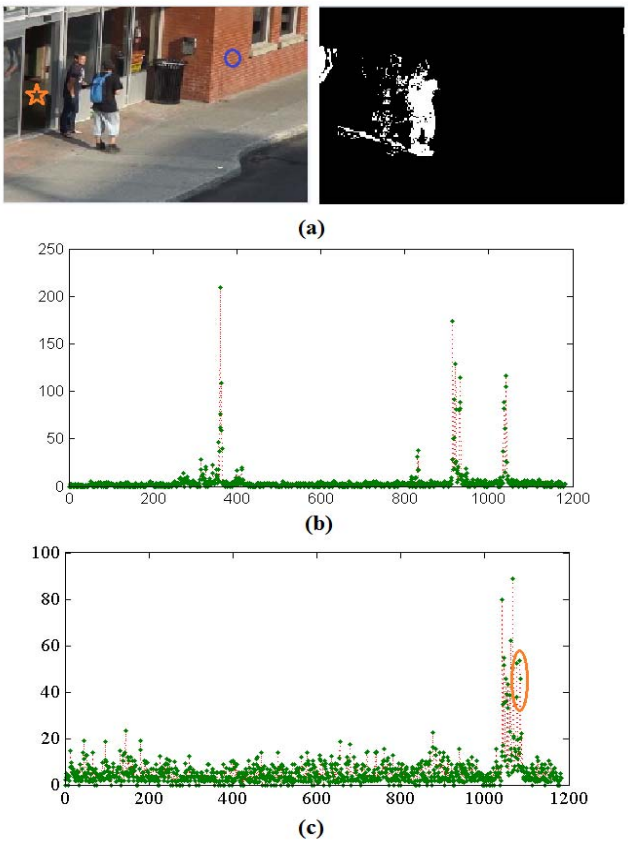

Fig. 8. Application of the proposed model to the Wall-flower bus-station dataset with (a) Frame 893 and BG-NARX evaluation of Frame 893 based on previous 5 frames, (b) ED/VA-based pixel alterations at pixel location (*) and (c) false positive induction due to abrupt shadow follow-up at pixel location (o)

C. Dataset three (Traffic scene under changing cloud cover)

Fig. 8 demonstrates the evaluation of BG-NARX model against video data taken from the MIT training dataset [17]. The dataset case presents almost all the challenges from the other scenarios including moving trees, dynamic shadows, intensity variations and variable-level pixel alterations due to vehicle/pedestrian movements. Similar to Dataset Two, the outcome presented a large number of falsely identified foreground pixels. Moreover, objects remaining static for extended periods, such as the vehicles waiting on signal seen on the right in Fig. 8 (a) were completely discarded. This does show a shortcoming of not using codebook's conventional maximum negative runtime length (MNRL) parameter which can be resolved by hybridizing the proposed technique with a "lightweight" codebook as implemented by Kim *et al*. [6] and Ilyas *et al*. [7].

IV. Evaluation of Benchmarking Datasets

The model was trained and evaluated over an Intel Win7 64-bit Core i7 machine with 8GB of RAM using Matab R2012b VS .Net 2012 (Framework 4.5) and OpenCV API. Despite the usage of EMGU CV wrapper API, the system offered real-time processing with Neural Training. To further establish the robustness of the proposed strategy, the algorithm was applied to objectively detect real objects from the noise induced in the background scenes. For each scene a connected-component labeling was performed based on manually specified pixel compactness and aspect ratios as shown in Table II:

TABLE II. Threshold criteria used for the detection of foreground objects in video frames (P = people, V=vehicles)

	Dataset I		Dataset II		Dataset III	
	P	V	P	V	P	V
Aspect Ratio	1	-	2.66	-	1	1
Compactness	0.85	-	0.66	-	0.6	0.6
P2S Ratio	0.3	-	0.05	-	0.007	0.007

In Table II, aspect ratio is calculated as a manually identified height-to-width ratio, compactness is the ratio of foreground-to-background pixel ratio within the bounding box of each eight-connected region and P2S ratio is the pixel-to-scene ratio of the foreground pixels in each bounding box against the resolution of the full image matrix. Each threshold value given in Table II was given a **±5%** variation to accommodate for minor structural differences. Identification accuracy along with false positive and false negative occurrence were calculated on all the sample frames based upon abrupt bounding-box losses or occurrences which were matched against manually labeled frame sequences. The identification outcomes for the three datasets are shown in Table III:

TABLE III. Identification accuracies, false positives and false negatives for the three benchmarking datasets (P = people, V=vehicles)

	Dataset I *35 frames*		Dataset II *894 frames*		Dataset III *150 frames*	
	P	V	P	V	P	V
Accuracy (samples) %	**33** *94.3*	-	**724** *81*	-	**97** *64.6*	**136** *90.66*
False Positives *samples*	*0*	-	*79*	-	*31*	*14*
False Negatives *samples*	*0*	-	*118*	-	*17*	*2*

Table III presented a high number of negatives primarily due to shadows and reflections induced on the tinted glass making about 13% of all labeled individuals present in 894 frames. Additionally there were also a substantial number of false positives (8.8%) mainly due to individuals staying static for extending periods of time. For Dataset III false positives values were relatively low for vehicles with approximately 9% and 1.3% for false negatives. False negatives for vehicles were again due to blockage from other movers or trees. However, the pedestrian accuracy for the MIT Dataset III showed a higher percentage (20.66%) of false positives. However, majority of these occurrences were spontaneous frames due to doors operated in the front building. False negatives generally happen due to individuals getting blocked due to other objects such as

poles or thick tree growth present at the bottom of the video sequence. It was envisaged that a suitably calibrated compactness threshold would further reduce this value.

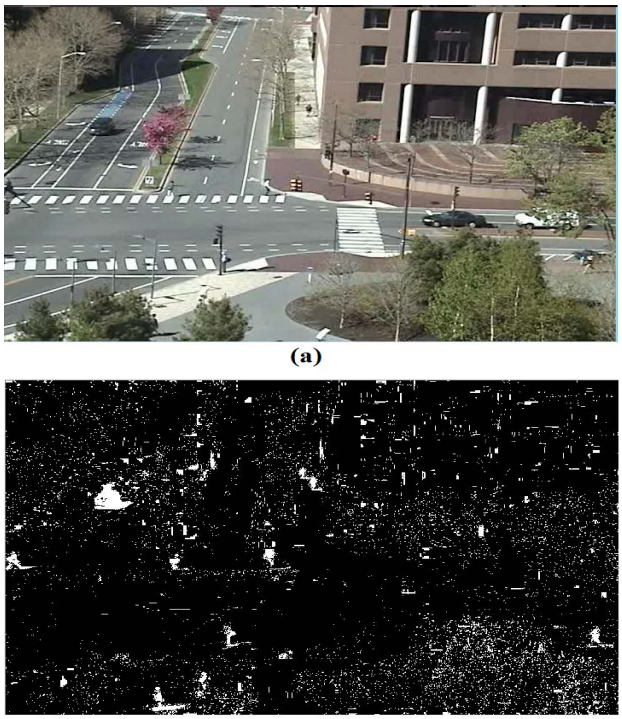

Fig. 8. Foreground/background modeling on (a) Original AVI dataset taken from MIT video database [17], (b) BG-NARX methodology

V. Conclusion and Future Direction

The paper reported a new technique to model foreground pixels in a foreground segmentation application via time series-based nonlinear dynamic neural networks to achieve background subtraction. The underlying rationale was based upon the differentiating ability of color and intensity variations in pixels as a function of angular and Euclidean-distance-based measures depicting the change behavior of a pixel over the last few frames. The core idea was based on the fact that, any overtime recurrent changes in a pixel's behavior due to dynamic backgrounds such as leaves or water fountains can be integrated into the background model based on a nonlinear neural training model. The outcomes were evaluated over a wide range of benchmarking datasets and presented promising outcomes. However, the outcomes showed a number of false-positives that were incorrectly classified as foreground pixels possibly because the alternating pixel-level behavior deviated from the normal routine for which the network was trained for. Majority of false positives occurred due to inherent shadows or reflections present within the scenes whereas the false negatives were generally attributed to foreground activity partly or completely occluded by other objects. As majority of these false identifications were spontaneous, it is envisaged that a discrete Bayesian or Hidden Markov Model based approach will help further improve bounding-box identification accuracy. Moreover, as a future prospect of this research, the abovementioned problem can be catered by training separate pixel-level or region-level neural networks. For instance, based on an area or pixel with a high MNRL, such as that occurring on roads for vehicular traffic, a long-delay neural network can be trained whereas for a low MNRL image region or pixel such as that occurring in tree/water-based areas, a smaller time-delay can be used.

References

[1] Wren, C.R.; Azarbayejani, A.; Darrell, T.; Pentland, A.P., Pfinder: Real-time tracking of the human body. Ieee Transactions on Pattern Analysis and Machine Intelligence, 1997. 19(7): p. 780-785.

[2] Haritaoglu, I.; D. Harwood; and L.S. Davis, W4: real-time surveillance of people and their activities. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2000. 22(8): p. 809-830.

[3] Stauffer, C.; Grimson, W. E. L.; "Adaptive background mixture models for real-time tracking," Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. , vol.2, no., pp.,252 Vol. 2, 1999

[4] Lee, D.-S.; Hull, J.J.; Erol, B., A Bayesian framework for Gaussian mixture background modeling. 2003 International Conference on Image Processing, Vol 3, Proceedings, 2003: p. 973-976.

[5] Harville, M., A framework for high-level feedback to adaptive, Per-Pixel, Mixture-Of-Gaussian background models. Computer Vision - Eccv 2002 Pt Iii, 2002. 2352: p. 543-560.

[6] Kim, K.; Chalidabhongse, T. H.;, Harwood, D; Davis L., Real-time foreground-background segmentation using codebook model. Real-Time Imaging, 2005. 11(3).

[7] Ilyas, A.; Scuturici, M.; Miguet, S., Real Time Foreground-Background Segmentation Using a Modified Codebook Model. Avss: 2009 6th Ieee International Conference on Advanced Video and Signal Based Surveillance, 2009: p. 454-459.

[8] Yusuf, A. S.; Wilkinson N.; Brown D. J. A Modified Codebook-based Background Subtraction Technique to improve Activity Classification in Highly Variable Environments. in International Conference on Industrial Engineering. 2012. Dubai, UAE: World Academy of Science, Engineering and Technology.

[9] He M.; DongSheng L.; Yongqin J.; Jianmin L., Forecasting Model on General Budget Revenue of Regional Finance Based on Dynamic Combination of BP Neural Network. in Information Science and Management Engineering (ISME), 2010 International Conference of. 2010.

[10] Lee, R. and Liu J., iJADE WeatherMAN: a weather forecasting system using intelligent multiagent-based fuzzy neuro network. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 2004. 34(3): p. 369-377.

[11] Rustagi, L.; Kumar L.; and Pillai G.N.. Human Gait Recognition Based on Dynamic and Static Features Using Generalized Regression Neural Network. in Machine Vision, 2009. ICMV '09. Second International Conference on. 2009.

[12] Junsong, Y.; Zicheng L.; and Ying W. Discriminative Video Pattern Search for Efficient Action Detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2011. 33(9): p. 1728-1743.

[13] Zhu, Q. and S. Zhan. Dynamic video segmentation via a novel recursive Bayesian learning method. in Image Processing (ICIP), 2010 17th IEEE International Conference on. 2010.

[14] Ahn, H.; Jae Joon A.; Kyong Joo, O. Facilitating cross-selling in a mobile telecom market to develop customer classification model based on hybrid data mining techniques. Expert Systems with Applications, 2011. 38(5): p. 5005-5012.

[15] Microsoft. Test Images for Wallflower Paper. 1999 [cited 2013 13th March]; Available from: http://research.microsoft.com/en-us/um/people/jckrumm/wallflower/testimages.htm.

[16] Toyama, K.; Krumm, J.; Brumitt, B.; Meyers, B. Wallflower: principles and practice of background maintenance. in Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on. 1999.

[17] Xiaogang, W.; Xiaoxu M.; and Grimson, W.E.L., Unsupervised Activity Perception in Crowded and Complicated Scenes Using

Hierarchical Bayesian Models. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2009. 31(3): p. 539-555.