AUTOMATIC KERNEL WIDTH SELECTION FOR NEURAL NETWORK BASED VIDEO OBJECT SEGMENTATION

Dubravko Culibrk
University of Novi Sad
Novi Sad, Serbia

Daniel Socek, Oge Marques, Borko Furht
Department of Computer Science and Engineering, Florida Atlantic University
Boca Raton FL 33431, USA

Keywords: Video processing, Object segmentation, Background modeling, BNN, Neural Networks.

Abstract: Background modelling Neural Networks (BNNs) represent an approach to motion based object segmentation in video sequences. BNNs are probabilistic classifiers with nonparametric, kernel-based estimation of the underlying probability density functions. The paper presents an enhancement of the methodology, introducing automatic estimation and adaptation of the kernel width.
The proposed enhancement eliminates the need to determine kernel width empirically. The selection of a kernel-width appropriate for the features used for segmentation is critical to achieving good segmentation results. The improvement makes the methodology easier to use and more adaptive, and facilitates the evaluation of the approach.

1 INTRODUCTION

Object segmentation is a basic task in the domain of digital video processing. Diverse applications, such as scene understanding, object-based video encoding, surveillance applications and 2D-to-pseudo-3D video conversion, rely on the ability to extract objects from video sequences.

The research into object segmentation for video sequences grabbed from a stationary camera has yielded a number of approaches based on the detection of the motion of objects. The approaches of this class scrutinize the changes observed between the consecutive frames of the sequence to detect pixels which correspond to moving objects. The task is particularly difficult when the segmentation is to be done for natural scenes where the background contains shadows, moving objects, and undergoes illumination changes.

For purposes of automated surveillance and scene understanding it is often of more interest not only to detect the objects moving in the scene, but to distinguish between two classes of objects:

- *Background objects* corresponding to all objects that are present in the scene, during the whole sequence or longer than a predefined period of time.
- *Foreground objects* representing all other objects appearing in the scene.

The goal of the foreground segmentation is to separate pixels corresponding to foreground from those corresponding to background.

Background Modeling Neural Network (BNN) represents a probabilistic approach to foreground segmentation(Culibrk et al., 2006). The network is a Bayes rule based unsupervised classifier designed to enable the classification of a single pixel as pertinent to foreground or background. A set of networks is used to classify all the pixels within the frame of the sequence.

The networks incorporate kernel-based estimators (Mood and Graybill, 1962) for probability density functions used to model the background. The networks are general in terms of the features of a pixel used to perform the classification, such as RGB values or intensity values. However, the accuracy of the process depends on the ability to determine the appropriate width for the estimator kernels. The process relies on empirical data and involves tedious experimentation. In addition the BNNs are unable to adapt to conditions occurring in specific sequences. Rather, a single value is typically used whenever the same features are used as basis for segmentation. Fur-

Culibrk D., Socek D., Marques O. and Furht B. (2007).
AUTOMATIC KERNEL WIDTH SELECTION FOR NEURAL NETWORK BASED VIDEO OBJECT SEGMENTATION.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - IU/MTSV*, pages 472-479

thermore, the complexity of the process of evaluation of a segmentation algorithm depends directly on the number of parameters that need to be specified (Wirth et al., 2006). Thus, reducing the number of parameters not only makes the methodology easier to use and more adaptive, it facilitates the evaluation of the approach.

In this paper, an extension of the BNN methodology is proposed to incorporate automatic selection of the appropriate kernel width. The proposed enhanced BNNs do not suffer from the above mentioned shortcomings of the fixed-kernel-width BNNs and achieve comparable segmentation performance.

The rest of the paper is organized as follows: Section 2 provides a survey of related published work. Section 3 contains a short description of BNNs. The proposed enhancement of the BNNs is described in Section 4. Section 5 is dedicated to the presentation of experimental evaluation results. Section 6 contains the conclusions.

2 RELATED WORK

Early foreground segmentation methods dealing with non-stationary background are based on a model of background created by applying some kind of low-pass filter on the background frames. The high-frequency changes in intensity or color of a pixel are filtered out using different filtering techniques such as Kalman filters (Karmann and von Brandt, 1990) to create an approximation of background in the form of an image (reference frame). The reference frame is updated with each new frame in the input sequence, and used to segment the foreground objects by subtracting the reference frame from the observed frame (Rosin, 1998). These methods are based on the most restrictive assumption that observe pixel changes due to the background are much slower than those due to the objects to be segmented. Therefore they are not particularly effective for sequences with high-frequency background changes, such as natural scene and outdoor sequences.

Probabilistic techniques achieve superior results in case of such complex-background sequences. These methods rely on an explicit probabilistic model of the background, and a decision framework allowing for foreground segmentation. A Gaussian-based statistical model whose parameters are recursively updated in order to follow gradual background changes within the video sequence is proposed in(Boult et al., 1999). More recently, Gaussian-based modelling was significantly improved by employing a Mixture of Gaussians (MoG) as a model for the probability density functions (PDFs) related to the distribution of pixel values. Multiple Gaussian distributions, usually 3-5, are used to approximate the PDFs (Ellis and Xu, 2001)(Stauffer and Grimson, 2000). The parameters of each Gaussian curve are updated with each observed pixel value. If an observed pixel value is within the 2.5 standard deviations (σ) from the mean (μ) of a Gaussian, the pixel value matches the Gaussian (Stauffer and Grimson, 2000). The parameters are updated only for Gaussians matching the observed pixel value, based on the following Equations:

$$\mu_t = (1-\rho) * \mu_{t-1} + \rho * X_t \tag{1}$$

$$\sigma_t^2 = (1-\rho) * \sigma_{t-1}^2 + \rho * (X_t - \mu_t)^T * (X_t - \mu_t) \tag{2}$$

where

$$\rho = \aleph(X_t, \mu_{t-1}, \sigma_{t-1}) \tag{3}$$

and $\aleph$ is a Gaussian function and X_t is a pixel value observed at time t. Equations 1 - 3 express a causal low-pass filter applied to the mean and variance of the Gaussian.

Using a small number of Gaussians leads to a rough approaximation of the PDFs involved. Due to this fact, MoG achieves weaker results for video sequences containing non-periodical background changes (e.g. due to waves and water surface illumination, cloud shadows, and similar phenomena), as was reported in (Li et al., 2004). The Gaussian-based models are parametric in the sense that they incorporate underlying assumptions about the probability density functions (PDFs) they are trying to estimate.

In 2003, Li et al. proposed a method for foreground object detection employing a Bayes decision framework (Li et al., 2004). The method has shown promising experimental object segmentation results even for the sequences containing complex variations and non-periodical movements in the background. The primary model of the background used by Li et al. is a background image obtained through low pass filtering. However, the authors use a probabilistic model for the pixel values detected as foreground through frame-differencing between the current frame and the reference background image. The probabilistic model is used to enhance the results of primary foreground detection. The probabilistic model is non-parametric since it does not impose any specific shape to the PDFs learned. However, for reasons of efficiency and improving results the authors applied binning of the features and assigned single probability to each bin, leading to a discrete representation of PDFs. The representation is equivalent to a kernel-based estimate with quadratic kernel. The width of the kernel used was determined empirically and remained fixed in all the reported experiments(Li et al., 2004). The system achieved performance better than that of the mixture of 5 Gaussians in the results presented in the same publication. However, when larger number of Gaussians is used, MoG achieved better performance (Culibrk, 2006). A nonparametric kernel density estimation framework for foreground segmentation and object tracking for visual surveillance has been proposed in (ElGammal et al., 2002). The authors present good qualitative results of the proposed system, but do not evaluate segmentation quantitatively nor do they compare their system with other methods. The framework is computationally intensive as the number of kernels corresponds to the number of observed pixel values. The width of the kernels is adaptive and they use the observed median of absolute differences between consecutive pixel values. The rationale for the use of median is the fact that its estimate is robust to small number of outliers. They assume Gaussian (normal) distribution for the differences and establish a relation between the estimated median and the width of the kernel:

$$\sigma = \frac{m}{0.68\sqrt{2}} \tag{4}$$

where m is the estimated median.

The approach based on background modelling neural networks was proposed in (Culibrk et al., 2006). The networks employ represent a biologically plausible implementation of Bayesian classifiers and nonparametric kernel based density estimators. The weights of the network store a model of background, which is continuously updated. The PDF estimates consist of fixed number of kernels, which have fixed width. The appropriate width of the kernels is determined empirically. The kernel width depends on the features used to achieve segmentation. Results superior to those of Li et al. and MoG with 30 Gaussians are reported in (Culibrk, 2006).The BNNs address the problem of computational complexity of the kernel based background models by exploiting the parallelism of neural networks.

3 BACKGROUND MODELING NEURAL NETWORK (BNN)

Background Modeling Neural Network (BNN) is a neural network classifier designed specifically for foreground segmentation in video sequences. The network is an unsupervised learner. It collects statistics related to the dynamic processes of pixel-feature-value changes. The learnt statistics are used to classify a pixel as pertinent to a foreground or background object in each frame of the sequence.

Note that a video sequence can be viewed as a set of pixel feature values changing in time, so-called *pixel processes* (Stauffer and Grimson, 2000). Pixel feature values are, in general, vectors in multi-dimensional space, such as RGB space. Probabilistic motion-based foreground segmentation methods, including the BNN approach, rely on a supposition derived from the definitions of foreground and background stated in Section 1:

Pixel (feature) values corresponding to background objects will occur most of the time, i.e. more often than those pertinent to the foreground.

Thus, if a classifier is able to effectively distinguish between the values occurring more frequently than others it should be able to achieve accurate segmentation. The BNN classifier strives to estimate the probability of a pixel value X occurring at the same time as the event of a background or foreground object being located at that particular pixel. In terms of probability theory, the BNN tries to estimate the joint probability $p(b,X)$ of background b and pixel value X occurring at the pixel the BNN is trying to classify and the analogous joint probability $p(f,X)$ for foreground.

By virtue of the Bayes rule, the classification criterion used in BNN is the following:

$$p(b|X)p(X) - p(f|X)p(X) > 0 \tag{5}$$

where $p(f|X)$ and $p(b|X)$ represent estimated conditional PDFs of an observed pixel value X indicating a foreground and background object, respectively. $p(X)$ is the estimated PDF of a feature value occurring.

In the structure of BNN, shown in Figure 1, three distinct subnets can be identified. The classification subnet is a central part of BNN concerned with approximating the PDF of pixel feature values belonging to background/foreground. It is a neural network implementation of a Parzen (kernel based) estimator (Parzen, 1962). The estimation is discussed in more detail in Section 4.

The classification subnet contains three layers of neurons. Input neurons of this network simply map the inputs of the network, which are the values of the feature vector for a specific pixel, to the pattern neurons. The output of the pattern neurons is a nonlinear function of Euclidean distance between the input of the network and the stored pattern for that specific neuron:

$$p_i = \exp[-\frac{(w_i - x_t)^T(w_i - x_t)}{2\sigma^2}] \tag{6}$$

where w_i is the vector of weights between the input neurons and the i-th pattern neuron, x_t is the pixel feature value vector and p_i is the output of the i-th

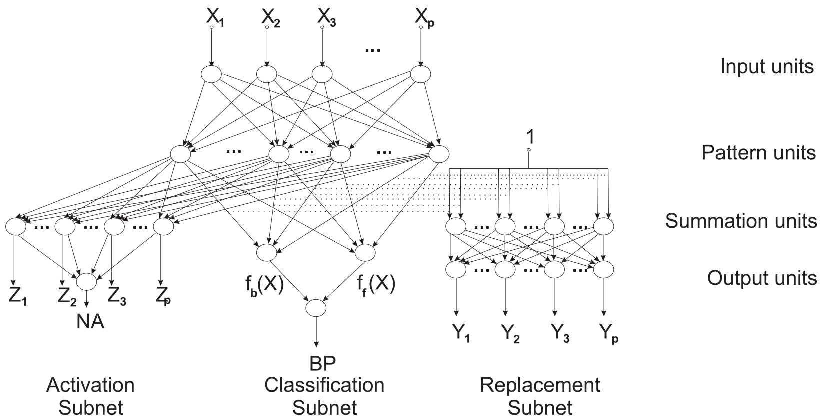


Figure 1: Structure of Background Modeling Neural Network.

pattern neuron. The only parameter of this subnet is the kernel width (σ), sometimes dubbed the smoothing parameter, which is used to control the shape of the nonlinear function.

The output of the summation units of the classification subnet is the sum of their inputs. The subnet has two summation neurons: one to calculate the probability of the observed pixel value corresponding to background and the other to calculate the probability of the value pertaining to foreground, corresponding to products in criterion 5.

Weights between the pattern and summation neurons are used to store the confidence with which a pattern belongs to the background/foreground. The weights of these connections are updated with each new pixel value received (i.e. with each frame), according to the following recursive equations:

$$w_{ib}^{t+1} = f_c((1-\frac{\beta}{P}) * w_{ib}^{t} + MA^t\beta) \qquad (7)$$

$$w_{if}^{t+1} = 1 - w_{ib}^{t+1} \qquad (8)$$

where w_{ib}^t is the value of the weight between the i-th pattern neuron and the background summation neuron at time t, w_{if}^t is the value of the weight between the i-th pattern neuron and the foreground summation neuron at time t, β is the learning rate, P is the number of the pattern neurons of BNN, f_c is a clipping function defined by (9) and MA^t indicates the neuron with the maximum response (activation potential) at frame t, according to (10).

$$f_c(x) = \begin{cases} 1, & x > 1 \\ x, & x \leq 1 \end{cases} \qquad (9)$$

$$MA^t = \begin{cases} 1, & \text{for neuron with maximum response;} \\ 0, & \text{otherwise.} \end{cases} \qquad (10)$$

Equations 7 - 10 express the notion that whenever an instance pertinent to a pattern neuron is encountered, the probability that that pattern neuron is activated by a feature value vector belonging to the background is increased. Naturally, if that is the case, the probability that the pattern neuron is excited by a pattern belonging to foreground is decreased. Vice versa, the more seldom a feature vector value corresponding to a pattern neuron is encountered the more likely it is that the patterns represented by it belong to foreground objects. By adjusting the learning rates, it is possible to control the speed of the learning process.

The output of the classification subnet indicates whether the output of the background summation neuron is higher than that of the foreground summation neuron, i.e. that it is more probable that the input feature value is due to a background object rather than a foreground object based on criterion 5. If the criterion 5 is satisfied than the pixel is classified as background, otherwise it is classified as foreground.

The activation and replacement subnets are Winner-Take-All (WTA) neural networks. The activation subnet performs a dual function: it determines which of the neurons of the network has the maximum activation (output) and whether that value exceeds a threshold (θ) provided as a parameter to the algorithm. If it does not, the BNN is considered inactive and replacement of a pattern neuron's weights with the values of the current input vector is required. If this is the case, the feature is considered to belong to a foreground object.

The first layer of this network has the structure of a 1LF-MAXNET (Kwan, 1992) network and a single neuron is used to indicate whether the network is active. The output of the neurons of the first layer of the network can be expressed in the form of the following equation:

$$Y_j = X_j \times \prod_{i=1}^{P} \{F(X_j - X_i | i \neq j)\} \quad (11)$$

where:

$$F(z) = \begin{cases} 1, & \text{if } z \geq 0; \\ 0, & \text{if } z < 0; \end{cases} \quad (12)$$

The output of the first layer of the activation subnet will differ from 0 only for the neurons with maximum activation and will be equal to the maximum activation. In Figure 1 these outputs are indicated with $Z_1,, Z_P$. Figure 2 shows the inner structure of a neuron in the first layer of the subnet. A single neuron in

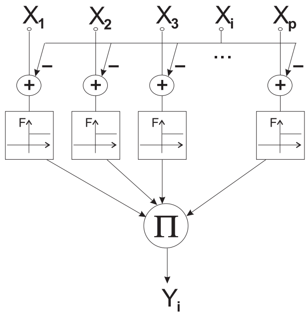


Figure 2: Structure of processing neurons of the activation subnet.

the second layer of the activation subnet is concerned with detecting whether the BNN is active or not and its function can be expressed in the form of the following equations:

$$NA = F(\sum_{i=1}^{P} Z_i - \theta) \quad (13)$$

where F is given by Equation 12 and θ is the activation threshold. Finally, the replacement subnet in Figure 1 can be viewed as a separate neural net with the unit input. Each of the first-layer neurons in the replacement subnet is connected with the input via synapses that have the same weight as the two output synapses between the pattern and summation neurons of the classification subnet. Each pattern neuron has a corresponding neuron in the replacement net. The function of the replacement net is to determine the pattern neuron that minimizes the criterion for replacement, expressed by the following equation:

$$replacement\_criterion = w_{if}^{t} + |w_{ib}^{t} - w_{if}^{T}| \quad (14)$$

The criterion is a mathematical expression of the idea that those patterns that are least likely to belong to the background and those that provide least confidence to make the decision should be eliminated from the model.

The neurons of the first layer calculate the negated value of the replacement criterion for the pattern neuron they correspond to. The second layer is a 1LF-MAXNET that yields non-zero output corresponding to the pattern neuron to be replaced.

To form a complete background-subtraction solution a single instance of a BNN is used to model the features at each pixel of the image.

4 AUTOMATIC KERNEL WIDTH ESTIMATION

BNNs employ a Parzen-estimator-based (Parzen, 1962) representation of the PDFs needed to achieve classification. This class of estimators is nowadays also known as kernel-based density estimators and was used in the approach presented in (ElGammal et al., 2002), as discussed in Section 2. A Parzen estimator of a PDF based on a set of measurements used within BNNs has the following analytical form:

$$p(X) = \frac{1}{(2\pi)^{\frac{P}{2}} T_o \sigma^P} \sum_{t=0}^{T_o} \exp[-\frac{(X - X_t)^T (X - X_t)}{2\sigma^2}] \quad (15)$$

where P is the dimension of the feature vector, T_o is the number of patterns used to estimate the PDF (observed pixel values), X_t are the pixel values observed up to the frame T_o, σ is the kernel width.

The Parzen estimator defined by (15) is a sum of multivariate Gaussian distributions centered at the observed pixel values. As the number of observed values approaches infinity, the Parzen estimator converges to its underlying parent density, provided that it is smooth and continuous. To reduce the memory and computational requirements of the segmentation, the BNNs employ a relatively small number of kernels (up to 30 kernels showed good results in our experiments), but the kernels are used to represent more than one observation and assigned weights in a manner similar to that discussed in (Specht, 1991). In addition, due to the format of the classification criterion 5 the normalizing expression preceding the sum in Equation 15 can be omitted for BNN purposes.

The smoothing parameter (σ) controls the width of the Gaussians. Fig. 3 shows the plot of a Parzen estimator for three stored points with values in a single dimension (e.g. if only the intensity value for a pixel is considered).

The horizontal lane in Fig. 3 represent the threshold values used to decide which feature values are

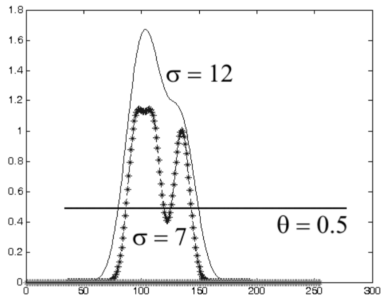


Figure 3: Plots of Parzen estimators for different values of ”smoothing parameter”.

covered by a single Gaussian. The threshold was set to 0.5 in the plot. All features within the circle defined by the cross-section of a single Parzen kernel and the threshold plane are deemed close enough to the center of the peak to be within the cluster pertinent to the Gaussian. The selection of smoothing parameter value and the threshold controls the size of the cluster. Larger values of σ lead to less pronounced peaks in the estimation, i.e. make the estimation ”smoother”.

The value of smoothing parameter has profound impact on the quality of segmentation and requires tedious experimentation to determine for a particular pixel feature used for segmentation. To alleviate this deficiency of the BNN approach, an automatic procedure is proposed for learning and adaptation of the smoothing parameter based on the properties of the segmented sequence.

The number of kernels in a BNN is fixed, and determined by the available computational resources. The smoothing parameter should be selected so that the predetermined number of kernels is able to cover all the pixel values occurring due to background. In addition, a more accurate approximation of PDFs can, in general, be achieved by the kernel of smaller width. Thus, the goal of smoothing parameter estimation is to determine the minimum width of kernels needed to account for all the background pixel values, based on a predefined number of kernels and the BNN activation threshold. To achieve this goal, the kernel width is updated with each new pixel value observed.

Let σ_i and μ_i be estimates of the standard deviation and mean of the background pixel values in the observed part of the video sequence, along the i-th dimension of the pixel feature vector. In order not to increase the computational and memory requirements of the BNN, it is desirable to estimate σ_i and μ_i based on the information already contained in the BNN.

For an estimate of the mean μ_i we use the average value of the i-th coordinate of patterns stored in the network, which correspond to the weights between the i-th input neuron and each pattern neuron of the classification subnet of BNN :

$$\mu_i = \frac{\sum_{j=1}^{P} w_{ij}}{P} \tag{16}$$

To estimate the standard deviation σ_i, the average of the distance of each center form μ_i is calculated, weighted with the weights of the connections between each pattern neuron and summation neuron corresponding to the background. This way the contribution of patterns likely to correspond to background is exacerbated, while the influence of the patterns due to foreground is diminished. The formula for σ_i is given by Equation 17.

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^{P} w_{bi} * (w_i j - \mu_i)^2}{P-1}} \tag{17}$$

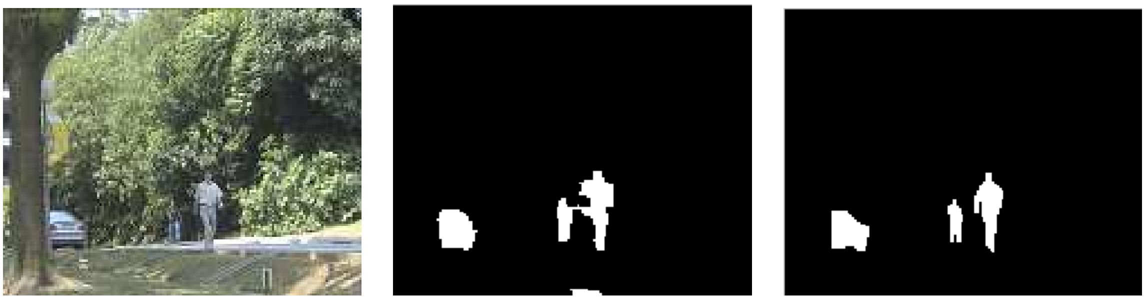
Since the width of the BNN kernels is the same along each dimension of the feature vector, maximum standard deviation over all the dimensions is used to estimate the smoothing parameter:

$$\sigma = 0.5 * \sqrt{\frac{-2 * max_{i \in 1..N_{dim}} \sigma_i}{\log \theta}} \tag{18}$$

where θ corresponds to the BNN activation threshold. Equation 18 corresponds to the kernel that will be active for all patterns within two estimated maximum standard deviations. The factor of two is introduced, since the estimator based on 17 tends to underestimate the real deviation. Equation 17 would give a precise estimate based on stored patterns, if all of them corresponded to background and the confidence of of them pertaining to background was one. It is unlikely that all the stored patterns in the BNN will correspond to background. In addition, the confidence that these patterns corresponds to the background will usually be less than one.

5 EXPERIMENTS AND RESULTS

To evaluate the approach, a PC-based foreground segmentation application based on the BNN methodology and employing adaptive kernels, has been developed. BNNs containing 30 processing neurons in the classification subnet were used. Pixel intensity was used as a feature to guide the segmentation. The activation threshold (θ) of the BNNs was set to 0.5, but the methodology showed itself to be robust to a wide range of threshold values. The learning rates used were 0.01, 0.005 and 0.003 depending on the dynamics of the sequence.



(a) (b) (c)

Figure 4: Results obtained for campus sequence: (a) frame from the original sequence, (b) segmentation results obtained for the frame shown, (c) ground truth frame.

(a) (b) (c)

Figure 5: Results obtained for room sequence: (a) frame from the original sequence, (b) segmentation results obtained for the frame shown, (c) ground truth frame.

A set of diverse sequences containing complex background conditions, provided by Li *et al.* (Li et al., 2004) and publicly available at `http://perception.i2r.a-star.edu.sg`, was used. The results of the segmentation were evaluated both qualitatively and quantitatively, using a set of ground truth frames provided by the same authors for the different sequences. The ten testing sequences were obtained in several different environments. They can be classified based on the sources of complexity in background variation, pertinent to each environment. Three classes of sequences (environments) can be identified: outdoor environments, small indoor environments and large (public) indoor environments.

The sources of complexity in the sequences obtained in outdoor environments are usually due to objects moved by wind (e.g. trees or waves) and illumination changes due to changes in cloud cover. For small indoor environments, such as offices, the source of complexity related mostly to objects such as curtains or fans moving in the background or screens flickering. The illumination changes are mostly due to switching lights on and off. Large public indoor environments (e.g. subway stations, airport halls, shopping centers etc.) are characterized by lighting distributed from the ceiling and presence of specular surfaces, inducing complex shadow and glare effects. In addition, these spaces can contain large moving objects such as escalators and elevators.

For reasons of space, frames illustrating qualitative results for a single sequence representative of each class are presented in this paper. More complete results as well as Matlab scripts that can be used to evaluate the approach in a manner similar to that described in (Culibrk, 2006) and (Li et al., 2004) can be found at `http://mlab.fau.edu`.

The frames shown in Figure 4 are pertinent to a video of a campus driveway. The complexity of the background in this sequence is due to the trees in the background moving violently in the wind and due to the changing illumination. The Figure shows the frame of the original sequence, segmentation result enhanced through morphological processing and human-generated ground truth. Figure 5 shows a segmentation result achieved for a small indoor environment sequence. The background is complex due to the moving curtain. Figure 6 shows a segmentation result achieved for a large indoor environment of a shopping mall. The changing glare of the floor makes the background complex in this case.

(a) (b) (c)

Figure 6: Results obtained for shopping mall sequence: (a) a frame of the original sequence, (b) segmentation result and (c) ground truth frame.

The results obtained for all three sequences and indeed all other testing sequences are good, although the intensity only based segmentation is prone to shadow segmentation effects as noted in (ElGammal et al., 2002).

6 CONCLUSION

Object segmentation is a fundamental task in several important domains of video processing. We proposed an enhancement of the Background modelling Neural Network approach to foreground segmentation. Automatic selection of the kernel width for the estimators used within the methodology, has been introduced. The proposed kernel width estimation principle enables the BNN to adapt to the conditions of a specific video, makes the methodology easier to use and allows for easier evaluation of the BNN approach.

The methodology has been tested using a publicly available and diverse set of sequences and achieves good segmentation results. Further testing of the methodology and application of the approach to problems of segmentation results enhancement and object tracking represent some of the possible directions of future research.

REFERENCES

Boult, T., Micheals, R., X.Gao, Lewis, P., Power, C., Yin, W., and Erkan, A. (1999). Frame-rate omnidirectional surveillance and tracking of camouflaged and occluded targets. In *Proc. of IEEE Workshop on Visual Surveillance, pp. 48-55*.

Culibrk, D. (2006). *Neural Network Approach to Bayesian Background Modeling for Video Object Segmentation*. Ph.D. Dissertation, Florida Atlantic University, USA.

Culibrk, D., Marques, O., Socek, D., Kalva, H., and Furht, B. (2006). A neural network approach to bayesian background modeling for video object segmentation. In *Proc. of the International Conference on Computer Vision Theory and Applications (VISAPP'06)*.

ElGammal, A., Duraiswami, R., Harwood, D., and Davis, L. (2002). Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. In *Proc. of the IEEE, vol. 90, No. 7, pp. 1151-1163*.

Ellis, T. and Xu, M. (2001). Object detection and tracking in an open and dynamic world. In *Proc. of the Second IEEE International Workshop on Performance Evaluation on Tracking and Surveillance (PETS'01)*.

Karmann, K. P. and von Brandt, A. (1990). Moving object recognition using an adaptive background memory. In *Timevarying Image Processing and Moving Object Recognition, 2, pp. 297-307*. Elsevier Publishers B.V.

Kwan, H. K. (1992). One-layer feedforward neural network fast maximum/minimum determination. In *Electronics Letters, pp. 1583-1585*.

Li, L., Huang, W., Gu, I., and Tian, Q. (2004). Statistical modeling of complex backgrounds for foreground object detection. In *IEEE Trans. Image Processing, vol. 13, pp. 1459-1472*.

Mood, A. M. and Graybill, F. A. (1962). *Introduction to the Theory of Statistics.* Macmillan.

Parzen, E. (1962). On estimation of a probability density function and mode. In *Ann. Math. Stat., Vol. 33, pp. 1065-1076*.

Rosin, L. (1998). Thresholding for change detection. In *Proc. of the Sixth International Conference on Computer Vision (ICCV'98)*.

Specht, D. F. (1991). A general regression neural network. In *IEEE Trans. Neural Networks, pp. 568-576*.

Stauffer, C. and Grimson, W. (2000). Learning patterns of activity using real-time tracking. In *IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, pp. 747-757*.

Wirth, M., Fraschini, M., Masek, M., and Bruynooghe, M. (2006). Performance evaluation in image processing. In *EURASIP Journal on Applied Signal Processing, Vol. 2006, pp. 13*.