Detection of Suspicious Pedestrian Behavior using Modified Probabilistic Neural Network

Tony Jan, Massimo Piccardi, and Thomas Hintz
Computer Vision Group,* University of Technology, Sydney
P.O. Box 123, Broadway, NSW, Australia

Abstract

In large scale visual surveillance applications, classification of human behaviors is very important. Classes of interest include suspicious human behaviors which should be effectively detected so as to alert supervisors' attention. In this paper, a data-based neural network such as the Modified Probabilistic Neural Network (MPNN) is introduced to approximately partition the classification space nonlinearly in order to achieve an acceptable classification performance while reducing computational complexity. The paper shows that this kind of network is able to achieve a good trade-off between classification accuracy and computational complexity. The performance of MPNN is compared to that of more conventional classification methods such as Hidden Markov Models (HMM) and the Multilayer Perceptron (MLP).

Keywords: human behavior recognition, neural network, pattern classification

1 Introduction

In many unmanned visual surveillance systems, automated detection of suspicious pedestrian behaviors is of particular importance in order to alert relevant authorities for attention [Foresti and Roli 2000], [Koller-Meier and Van Gool 2001]. Detection of suspicious human behavior involves the modeling and classification of human behaviors with certain rules. However, the modeling of human behavior is nontrivial since the observed input space of human movement can be very large due to the apparent randomness and complexity in human behavior. The idea is to partition this observed input space into the discrete states of the human movements and then classify the human behavior appropriately.

One possible approach is to use the state-space based modeling [Bobick and Wilson 1995], [Campbell and Bobick 1995], [Starner and Pentland 1995]. The state-space approach defines each static posture (position) as a state and describes a motion sequence by the composition of these states with some transitional probabilities [Goddard 1994]. For activity recognition, the joint probability is calculated through the states (motion sequence), and then the most likely motion sequence is selected for classification. Some well known examples of this method are Hidden Markov Model (HMM) [Yamato et al. 1992] and its variations [Brand 1997]. These methods are based on well-established mathematical frameworks. However, their implementation is difficult due to their intrinsic nonlinearity which generally does not provide closed-form solutions. Some approximation methods [Hammersley 1964], [Pavlovic and Rehg 2000] such as dynamic programming, variational inference, and other techniques have hence been proposed.

*email:jant@it.uts.edu.au

Another approach is based on Artificial Neural Network (ANN) techniques such as the Multilayer Perceptron (MLP) [Haykin 1996], [Rumelhart et al. 1986] or Self Organising Map (SOM) [Kohonen 1995]. These stateless data-modeling methods can be trained heuristically to nonlinearly partition the input space. However, their use is difficult due to large computations and intractability. Nevertheless, the nonlinear partitioning is necessary for many complex modelling and classification, and it is feasible that some approximated nonlinear partitioning could still achieve acceptable classification while reducing computational complexity to a reasonable level. This is an engineering solution where one step back in output performance would guarantee several steps ahead in overall performance including reduced computational complexity and possible recovery of tractability.

In this paper, the Modified Probabilistic Neural Network (MPNN) which approximates the General Regression Neural Network (GRNN) is introduced to classify the human behavior in a car park as either suspicious or unsuspicious. The GRNN algorithm shows its nearest optimality in Bayesian sense and has proven to work very well for classification. The paper examines the performance of MPNN and compares its performance to other conventional methods such as HMM and MLP.

2 Method

In wide-area visual surveillance applications, human "behaviors" are often interpreted from human movement. In this work, the detection of suspicious behavior follows the steps reported hereafter. First, moving objects in the scene are detected by subtracting a model of the background scene from the current frame. The background model and subtraction are computed in accordance with the approach presented in [Cucchiara et al. 2000]; however, other background subtraction methods could be successfully exploited, or, alternatively, other techniques based on the optical flow analysis or feature tracking.

The moving objects detected as a result of the background subtraction are typically pedestrians and vehicles, and combination of them such as groups of people. Since our goal is the classification of pedestrians' behaviors, relevant features from those subjects must be extracted and reliably tracked. To this aim, we assume that trajectory-related information such as the velocity of subjects' head or body movement at each time frame can provide adequate information for behavior classification.

In particular, in this work we decided to use subjects' head as feature since heads can be rather accurately extracted from the detected moving objects. The technique used is a simple template matching under weak perspective assumptions. An example of successful head extraction process is shown in Fig. 3 to Fig. 6.

When different sets of movement information are displayed in the observed input vector space, the constellation of different behaviors form distinct clusters. One or more clusters will correspond to the suspicious behavior and should be used to generate an output to alert the security person. However, the clustering (partitioning) of these behaviors is a nontrivial task. In the next section, the MPNN is introduced which can partition the input space effectively with reduced computations.

2.1 Introduction to MPNN

The MPNN was initially introduced by Zaknich et al in [1991]. It is closely related to Specht's GRNN and his previous work, Probabilistic Neural Network (PNN) [Specht 1990]. The basic MPNN and GRNN methods have similarities with the method of Moody and Darken [1990]; the method of RBF's [Powell 1974], [Broomhead and Lowe 1988]; the CMAC [Miller et al. 1990]; and a number of other non-parametric kernel-based regression techniques stemming from the work of Nadaraya and Watson [1964].

A standard version of the GRNN equation, which is similar to the Nadaraya and Watson equations, is

$$\widehat{y}(\underline{x}) = \frac{\sum_{i=1}^{\mathcal{NV}} y_i exp\frac{-(\underline{x}-\underline{x}_i)^T(\underline{x}-\underline{x}_i)}{2\sigma^2}}{\sum_{i=1}^{\mathcal{NV}} exp\frac{-(\underline{x}-\underline{x}_i)^T(\underline{x}-\underline{x}_i)}{2\sigma^2}} \tag{1}$$

where

$\underline{x}$= training vector for class i in input space;
$\underline{x}_i$= single training vector in the input space
σ= single learning or smoothing parameter
y_i= scalar output related to $\underline{x}_i$
NV= total number of training vectors

In above GRNN equation 1, each and every training data pair$\{\underline{x}_i\text{->}y_i\}$ is incorporated into its architecture, ($\underline{x}_i$ is a single training vector in the input space, and y_i is the associated desired scalar output). This requires very large computations.

If it can be assumed that there is a corresponding scalar output y_i for each local region of the input space which is represented by a centre vector $\underline{c}_i$, then the general algorithm of MPNN given in equation 2 can nonlinearly approximate GRNN equation within acceptable accuracy. The centre vectors $\underline{c}_i$ for each cluster can be readily estimated from K-means clustering algorithms [Zaknich et al. 1991].

This method reduces the complexity in computation significantly while performing acceptable partitioning for classification. The general algorithm for the MPNN is then:

$$\widehat{y}(\underline{x}) = \frac{\sum_{i=o}^{M} Z_i y_i f_i(\underline{x})}{\sum_{i=0}^{M} Z_i f_i(\underline{x})} \tag{2}$$

with

$$f_i(\underline{x}) = exp\frac{-(\underline{x}-c_i)^T(\underline{x}-c_i)}{2\sigma^2} \tag{3}$$

where

$\underline{x}$= input vector
$\underline{c}_i$ = center vector for class i in the input space
σ= learning parameter
y_i= output related to $\underline{c}_i$
Z_i= no. of vectors $\underline{x}_i$associated with each $\underline{c}_i$
M = number of unique centers $\underline{c}_i$

A Gaussian function is often used for $f_i(\underline{x})$ as defined in equation 3. However, many other suitable radial basis functions can be used. Tuning simply involves finding the optimal σ giving the minimum mean squared error (mse) of the network output minus the desired output for a representative tuning set of known sample vector pairs by a convergent optimization algorithm. [Zaknich et al. 1991]

3 Experiment

3.1 Overview of experiment

As an experiment, a number of different user scenarios in a parking space were simulated. In each scenario, an actor acted in either a normal or suspicious behavior. In normal behaviors, the actor entered the parking area and walked to their car with natural movements. In suspicious behaviors, a person kept wandering around cars, with possible intention to steal belongings or damage a car. Each scenario lasted for 10 seconds, with frames sampled at 5 Hz, thus producing 50 frames per scenario. The starting time of each scenario was chosen randomly, assuming that a behavior would last for much longer than the scenario's duration.

Figure 1: Surveillance Image from Open Car Park

For each scenario, the subject's head was located and tracked. The magnitude of the head velocity at each frame was then computed and stored for behavior classification. In this way, each scenario provided a 49-feature vector describing the pedestrian's speed pattern. Figure 1 shows an example of typical normal and suspicious pedestrian behavior.

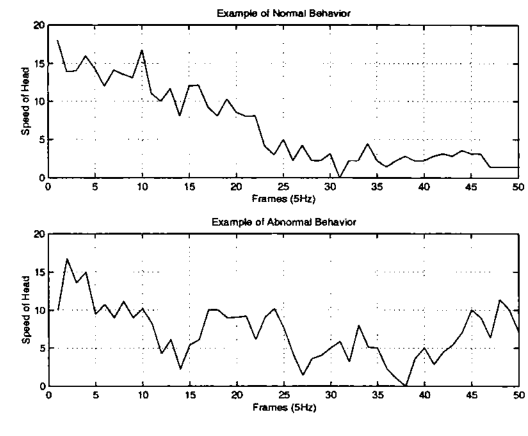

Figure 2: Speed of head movement at each time frame for normal pedestrian behavior (above) and suspicious pedestrian behavior (below)

In this paper, 160 samples from the normal and suspicious behaviors each were used to train the network. It should be noted that 160 samples may not be considered a large training data size. We intentionally limited the training sample in order to see the network performance in limited training sample conditions. For performance comparison, other 72 independent user behavior samples were classified and the performance was compared.

3.2 Performance

3.2.1 Modified Probabilistic Neural Network

The MPNN had 20 input states and Gaussian Radial Basis Functions for nonlinear approximation. The head's speed was quantized to 20 states (as shown in Figure 2-vertical axis) and the number of input states were chosen accordingly. Section 2.1 described the network architecture in detail. The classification performance of MPNN is reported in Table 1.

Input\Classified	Normal	Suspicious
Normal behavior	98%	2%
Suspicious behavior	4%	96%

Table 1: Performance table for MPNN

In this kind of application, the most serious type of error occurs when a suspicious behavior is misclassified as normal behavior. This corresponds to false dismissals of dangerous situations. Consequently, an optimal classification should reduce this error as much as possible. For MPNN, this error is limited to 4% only. This means that 96% of pedestrians who show suspicious behavior will be detected correctly and the security officer will be alerted accordingly.

3.2.2 Hidden Markov Model

The HMM network was a left to right network with 10 hidden states and 20 observation states. The number of observation states was set at 20 to model 20 quantized head speed values, and the number of hidden states was chosen empirically. A higher hidden state number resulted in a too sparse state space and a lower number resulted in compromised accuracy. Learning was performed by standard backward-forward technique with Expectation Maximization algorithms [Rabiner 1989].

The classification performance was:

Input\Classified	Normal	Suspicious
Normal behavior	89%	11%
Suspicious behavior	28%	72%

Table 2: Performance table for HMM

Results in Table 2 shows that 28% of suspicious behaviors were misclassified as normal behavior and that only 72% of suspicious behavior alerted the security officer correctly. The poor performance of HMM in this experiment is likely due to the lack of training data that represents all possible behaviors, or the limited number of features used for classification. This again emphasizes the importance of a network that can interpolate within the given training sample in order to learn all of possible input space effectively.

3.2.3 MultiLayer Perceptron

The MLP network had 1 input layer with 20 neurons, 1 output layer with 2 possible outputs, and 1 hidden layer with 5 hidden neurons. The number of neurons in the input layer corresponds to possible head speeds and other parameters were chosen empirically. The training was performed by the standard back-propagation algorithms. The classification performance was:

Input\Classified	Normal	Suspicious
Normal behavior	97%	3%
Suspicious behavior	8%	92%

Table 3: Performance table for MLP

The performance of MLP was very competitive in terms of accuracy. Also the time required at run time to perform the classification of a scenario is reasonably limited and definitely less than the scenario duration itself. The MLP only drawback might be its long training time (in the order of a few hundreds of seconds with a normal PC), which is expected to grow more than linearly with the number of examples. This long training time may affect widespread use in real life applications where the training set should be considerably larger than in our experiment.

3.3 Results and Analysis

The results show that MPNN was able to achieve more accurate classification in this experiment where the training data are limited. As the training data increase, it is possible to show that the accuracy of HMM and MPNN becomes comparable. However, the training time increases significantly for both. This confirms that the MPNN can provide acceptable classification accuracy with less computation, since it was proven able to achieve good classification with only limited training data.

MLP classification provided comparable accuracy to MPNN for the same training data. However, its training time is expected to be exceedingly long in many real life visual surveillance applications with a substantially larger training set.

4 Conclusion and Future Works

In this paper, it was shown that an MPNN is able to perform effective classification in a visual surveillance application, even with a limited training data set. In the experiments performed, MPNN proved superior to other methods such as HMM and MLP.

A possible future approach could be that of combining techniques based on ANNs with HMM: the nonlinear classification capacity of ANNs combined with the probabilistic inference framework of HMM might be able to achieve better performance than either separate approach. Some works have already been proposed on combining MLP with HMM for speech recognition [Xinye et al. 2000]. However, MPNN combined with HMM seems to promise improvement over MLP-HMM network in terms of effectiveness and computational efficiency.

The clustering of MPNN can further be improved using various clustering algorithms such as Kohonen's Self Organising Map [Kohonen 1995] or popular Support Vector Machine (SVM) [Burges 1998] methods.

Figure 3: Current Frame of the Scene at time t

Figure 4: Background Estimation at time t

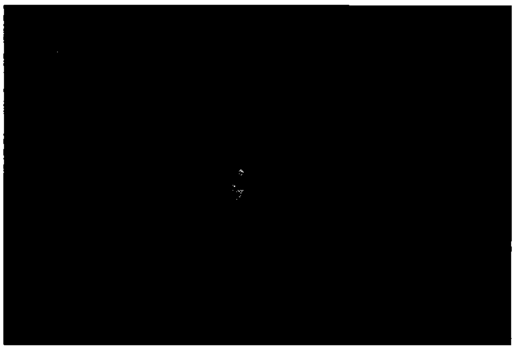

Figure 5: Subject Detected in the Current Frame at time t

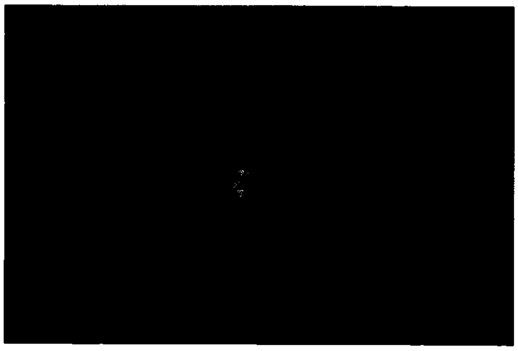

Figure 6: Subject Head Detected in the Current Frame at time t

References

FORESTI, G. L. AND ROLI, F. 2000. Real-time recognition of suspicious events for advanced visual-based surveillance. In *Multimedia Video-Based Surveillance Systems:Requirements, Issues and Solutions*, Kluwer Academic Publishers, 2000, pp. 84-93.

KOLLER-MEIER, E.B. AND VAN GOOL, L. 2001. Modeling and recognition of human actions using a stochastic approach. In *Proceedings of 2nd European Workshop on Advanced Video-Based Surveillance Systems*, September 2001, pp. 17-28.

YAMATO, J., OHYA, J., AND ISHII, K. 1992. Recognizing human action in time sequential images using Hidden Markov Model. In *Proc. IEEE Conf. CVPR*, pp. 379-385, Champaign, IL, 1992.

GODDARD, N. H. 1994. Incremental model-based discrimination of articulated movements from motion features. In *Proc. of IEEE Computer Society Workshop on Motion of Non-Rigid and Articulated Objects*, pp. 89-95, Austin, TX, 1994.

BOBIK, A. F., AND WILSON, A.D. 1995. A state-based techniques for the summarization and recognition of gesture. In *Proc. of 5th Intl. Conf. on Computer Vision*, pp. 382-388, 1995.

NADARAY, E. A. 1964. On estimating regression. *Theory Probab. Appl.*, vol. 9, pp. 141-142, 1964.

CAMPBELL, L. W. AND BOBICK, A. F. 1995. Recognition of human body motion using phase space constraints. In *Proc. of 5th Intl. Conf. on Computer Vision*, pp. 624-630, 1995.

STARNER, T. AND PENTLAND, A. 1995. Visual recognition of American Sign Language using Hidden Markov Models. In *Proc. Intl. Workshop on Automatic Face and Gesture Recognition*, Zurich, Switzerland, pp. 189-194, 1995.

BRAND, M. 1997. Coupled Hidden Markov models for modeling interacting processes. *MIT Media lab technical report* 405, June 1997.

KOHONEN, T. 1995. *Self-Organising Maps*, Springer-Verlag, 1995.

HAYKIN, S. 1996. *Neural Networks*, Prentice Hall, 1996

RUMELHART, D. E., HINTON, D. E., AND WILLIAMS, R. J. 1986. Learning representation by backpropagation errors. *Nature*, vol. 323, no. 9, pp. 533-536, 1986.

WERBOS, P. J. 1974. *Beyond regression: New tools for prediction and analysis in the behavioral sciences.* PhD dissertation, Comm. Appl. Math,. Harvard Univ., Cambridge, MA, Nov. 1974.

MOODY, J. AND DARKEN, C. 1989. Fast learning in networks of locally tuned processing units. *Neural Comput.*, vol. 1, no. 2, pp. 281-294, 1989.

MOODY, J. AND DARKEN, C. 1988. Learning localized receptive fields. In *Proc. 1988 Connectionist Models Summer Sch.*, Touretzky, Hinton, and Sejnowski, Eds. San Francisco, CA: Morgan-Kaufmann, 1988, pp. 133-143.

POWELL, M. J. D. 1974. *Radial basis functions for multivariate interpolation: A review.* Tech. Rept. DAMPT 1985/NA 12 Dept. Appl. Math. Theoretical Phys. Cambridge, MA, Nov. 1974.

BROOMHEAD, D. S. AND LOWE, D. 1988. Radial basis functions, multi-variable functional interpolation and adaptive networks. *Royal Signals Radar Est. Memo. 4248*, Mar. 28, 1988.

ALBUS, J. S. 1975. A new approach to manipulator control: The cerebellar model articulation controller (CMAC). *J. Dyn. Syst. Meas. Contr.*, Sept. 1975, pp. 220-227.

MILLER, W. T., GLANTZ, F. H., and KRAFT, L. G. 1990. CMAC: An association neural network alternative to backpropagation. *Proc. IEEE*, vol. 78, pp. 1561-1567, Oct. 1990.

SPECHT, D. F. 1990. Probabilistic Neural Networks. *Int. Neural Network Soc., Neural Networks*, vol. 3, pp. 109-118, 1990.

HAMMERSLEY, J. M. AND HANDSCOOMB, D. C. 1964. *Monte Carlo Methods*, London: Chapman and Hall, 1964.

NADARAYA, E. A. 1964. On Estimating Regression. *Theo. Probab. Appl.*, vol. 9, pp. 141-142.

PAVLOVIC, V. AND REHG, J. M. 2000. Impact of dynamic model learning on classification of human motion. *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on* , Vol. 1 , 2000, pp. 788 -795.

RABINER, L. R. 1989. A Tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, vol. 77, No. 2, Feb. 1989.

Burges, C. J. C. 1998. A Tutorial on Support Vector Machine for Patter Recognition. *Kluwer Academic Publisher*, Boston.

XINYE, H., M. XIAOHUI, L. XIANG, F. YUQING, L. AND JIREN, L. 2000. Study on recognition of speech based on HMM/MLP hybrid network. *Proceedings of 5th International Conf. on Signal Processing, 2000, WCCC-ICSP 2000*, vol. 2, pp. 718-721.

ZAKNICH, A., DESILVA, C. J. S., AND ATTIKIOUZEL, Y. 1991. Modified Probabilistic Neural Network for Nonlinear Time Series Analysis. *Neural Networks 1991, International Joint Conference On*, pp. 1530-1535, vol. 2, 1991.

CUCCHIARA, R., GRANA, C., PICCARDI, M., AND PRATI, A. 2000. Statistic and knowledge-based moving object detection in traffic scenes *Proceedings of the 2000 IEEE Intelligent Transportation Systems Conference*, pp. 27 -32.

LIPTON, A. J., FUJIYOSHI, H., AND PATIL, R. S. 1998. Moving target classification and tracking from real-time video *Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision (WACV '98)* , 1998, pp. 8 -14.