Trajectory classification based on Hidden Markov Models

Jozef Mlích
xmlich02@stud.fit.vutbr.cz
Department of Computer Graphics and Multimedia
Brno University of Technology

Petr Chmelař
chmelarp@fit.vutbr.cz
Department of Information Systems
Brno University of Technology

Abstract

This paper presents a method for statistical modeling and classification of motion trajectories using Hidden Markov Models. Mass recordings from visual surveillance are processed to extract objects trajectories. Hidden Markov Models of classes of behaviour are created upon some annotated trajectories. In this way, information about complex object behaviour of objects can be discovered.

Additionally, an experiment shows the successful application of Hidden Markov Models on trajectories of people in an underground station in Roma. Finally, a comparison of efficiency on different data sets, is discussed.

Keywords: Surveillance, Hidden Markov Models, Trajectory classification

1. INTRODUCTION

Nowadays, large amount of data is produced by a constantly increasing number of surveillance cameras. This data is a potential source of useful information; however, it is difficult to obtain the information in general. Suitable processing methods are often not available. In the face of present day security threats, it is widely demanded to apply computer vision and machine learning methods on surveillance data.

It is advantageous to process the mass recordings from surveillance cameras using computer vision techniques. The common way is to process video sequence as separated frames. In this way there can be analysed colour, texture and some structure features of the scene. In this way, an information about complex object behaviour cannot be extracted, because the relation between frames is not present. The presented task, on the contrary, assumes that it is possible to obtain the temporal and motion analysis and to identify individual objects in the scene. Other techniques are used to identify objects and its motion parameters. For instance it is background subtraction, optical flow, motion estimation and segmentation. Kalman and particle filters are widely supposed to be the best solution of the tracking problem [Intel 2007].

Although this task is considered to be hard, especially in crowded scenes, we think it is possible to discover security threats and to deduce information about complex behaviour of objects from its trajectory. However, this knowledge is even more hard to gain. It is similar to the other hi-level feature detection task as face or gait recognition. For that purpose the machine-learning techniques appears to be necessary. There are wide-spread discriminative classification techniques based on Bayesian and kernel methods or neural networks describing and separating classes of objects based on their description [Bishop 2006]. However, these techniques are not very suitable for classification of the variable length of the data sequences the trajectory is supposed to be.

The trajectory is the path a moving object follows through space. It is represented as a sequence of spatio-temporal points. The need of knowledge discovery in the trajectory data leads to the linear dynamic and Markov models for the data classification [Roweis and Ghahramani 1997]. The presented approach is based on supervised learning and classification using Hidden Markov Models.

The common experience suggests that the majority of trajectories from surveillance cameras in an underground station belong to normal “boring” situations. Contrary, many abnormal and potentially interesting scenarios exist and we are about to present a solution. We suppose, that is possible to identify situation that may affect single person in recordings like is a personal injury, turnpikes jumping or vandalism. Next we suppose, that it will be possible to identify situations involving couple of people. Example of such situations could be a fight or other kind of violence. There might be also situations that may affect all or almost all people; for example gun shots, underground fire or a train accident. And, the most challenging task is to identify situations like pickpocketing or banditry, which still seems to be impossible.

To meet these challenges, the paper presents theoretical background of Hidden Markov Models together with the related works in the following chapter. The description of the created system is in the chapter three. The experiment is described in chapter four and evaluated in the fifth chapter. The last chapter summarizes the work, achieved results and proposes further topics of research.

2. BACKGROUND

Hidden Markov Models (HMM) approach belongs to supervised learning and statistical modelling methods for sequential data [Bishop 2006]. It has been used prominently and successfully in speech recognition and, more recently, in handwriting recognition and Visual recognition of Sign Language [Starner 1995]. Example of an HMM is shown in Figure 1. The sample model is described as a graph with four internal and two marginal states connected by (oriented) transitions [Černocký 2003]. Moreover, there are six output vectors associated in Figure 1.

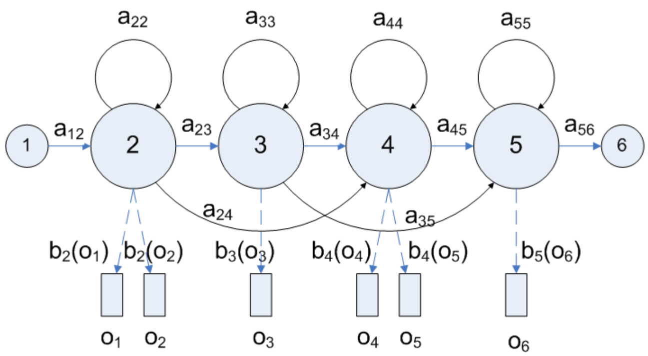

Figure 1: *Example configuration of a Hidden Markov Model*

The trajectory classification is similar to the speech recognition tasks. A trajectory is a continuous quantity, that can be described analytically as the position of the object in time. For common reasons, its discrete representation is used together with its temporal derivations (velocity and acceleration). Thus, an object trajectory O is a potentially infinite sequence of state vectors $o(t) = [x, y, d_x, d_y, d_x^2, d_y^2]$.

The trajectory classification problem can be formulated as to identify the class $c_i(i = 1..N)$ to which belongs the trajectory state sequence. The basic formulation of the problem is given by maximization of a conditional probability:

$$i^* = arg \max_i P(c_i|O) = arg \max_i \frac{P(O|c_i)P(c_i)}{P(O)} \quad (1)$$

We use Bayes theorem in (1), because we cannot evaluate $P(c_i|O)$ directly. Assuming we know prior probabilities $P(c_i)$ and $P(O)$, we are about to compute the likelihood $P(O|c_i)$; the probability of the sequence O knowing the class c_i. To compute this, we should have a model M for class c_i. The model is a finite state automaton with K states generating sequence O. There are transition probabilities $a_{k,j}$ between the states. Except first and the last state, states are emitting or generating output probability density function $b_j(o(t))$, as illustrated in Figure 1.

In the figure, there is a sample configuration of $A = [a_{k,j}]$ $(k, j = 1..K)$, the transition matrix, which defines the probability of transition to the next state for each combination of HMM states. The example is taken from [Černocký 2003]. The corresponding sample HMM sequence or path through the model is $X = \{1, 2, 2, 3, 4, 4, 5, 6\}$. However, this information is from the view of the trajectory state sequence hidden. The probability of passing an object O through a model M by a way X. is defined by equation 2.

$$P(O, X|M) = a_{x(o)x(1)} \prod_{t=1}^{T} b_{x(t)}(o_t)\, a_{x(t)x(t+1)} \quad (2)$$

Viterbi algorithm [Young et al. 2006] (defined by Equation 3) finds the most probable way through the model. The algorithm is used to evaluate the model by maximizing probability of correspondence with a trajectory class.

$$P^*(O|M) = \max_{\{X\}} P(O, X|M) \quad (3)$$

For training the model M_i, corresponding to the trajectory class c_i, the Baum-Welch algorithm [Young et al. 2006] is used. It is a generalized expectation-maximization algorithm defined by Equation 4 that modifies weights of transitions and statistics of the models.

$$P(O|M) = \sum_{\{X\}} P(O, X|M) \quad (4)$$

For more detailed information on training HMM see [Bishop 2006], [Young et al. 2006] or [Roweis and Ghahramani 1997].

2.1. Related works

Most of the related papers deal with the motion analysis, tracking and recognizing human activities altogether [Ohya et al. 2005], [Chellappa et al. 2005], [Aggarwal and Cai 1999], [Stauffer and Grimson 2000]. However, we deal only with the behaviour discovery from spatio-temporal sequence of points extracted using [Intel 2007], the other options to identify abnormal behaviour in surveillance recordings could be realised for example by analysis of dynamic texture using optical flow [Chan and Vasconcelos 2005], although it is impossible to identify individual behaviour of objects in this way.

There are several approaches of statistical modelling and classification of motion trajectories in the literature. The survey of proposed methods is described in [Aggarwal and Cai 1999] and above mentioned. These papers deal with template matching, where nearest neighbour search to the predefined models is used or state-space approaches such as fitting a curve or some application of Hidden Markov Models. However, we haven't found there any exact evaluation of such methods in video surveillance of public places.

3. SYSTEM OVERVIEW

The process of classification is illustrated in Figure 2. Because tracking is a difficult task and the output of a simple tracker even using Kalman filter [Intel 2007] is usually very noisy. It is good to make some pre-processing to improve the accuracy of the input trajectories.

The exploitation of HMM classifier is divided in two phases in the proposed system – the training and classification phase.

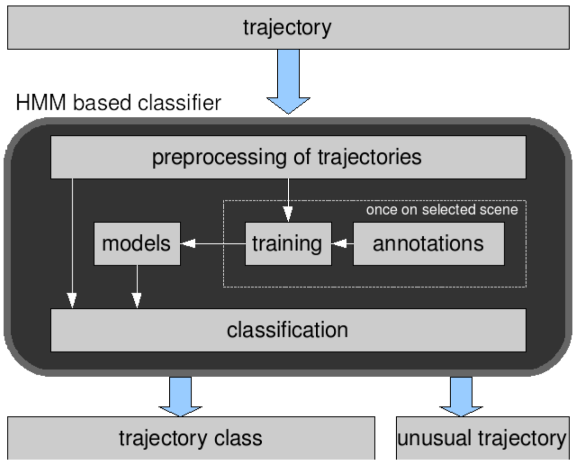

Figure 2: *Schema of trajectory processing.*

The training phase is in the schematic diagram (Figure 2) highlighted using a dashed box and it is done once for each camera. First, we define all normal trajectory classes in the scene. Secondly, initial models are defined for each trajectory class. This part is done by user. Next, the trajectory classes in video sequence are marked (annotations) and models are adapted according to them (training process). The inputs for this phase are the initial model and the annotated trajectories, the output is a new classification model. For better model estimation in the training phase is more suitable to use the most representative trajectories.

The classification step produces evaluation for each trajectory. This part compares a trajectory with all models and decides which model best fits the trajectory. The inputs of the classification step are trajectories and classification models. The results are coefficients, which express the degree of correspondence of the trajectory with each class. According to the degree of correspondence, the class which best fits the trajectory is determined. In case the likelihood of trajectory is low for all models, the trajectory is assumed to be abnormal. The threshold for abnormal trajectories must be set by hand.

A difficult question is how to choose the initial model for the training. The simplest approach is to setup by hand a few models and to choose initial model according to the best degree of correspondence for the selected scene. Whereas, very simple model could not affect a complex trajectory and a complex model needs much more data for its training. The topology of the initial model is shown in Figure 1.

4. EXPERIMENTS

For a successful classification of trajectories and for search of abnormal trajectories, it is important to have a well defined scene with well defined scenarios, because anomalous trajectories correspond to scenarios that are not defined. It is very difficult to define all possible normal scenarios in all scenes; therefore we select scenes with simpler scenarios.

For example, a scene with a platform, ticket machine and five gates is a complicated one, because it is necessary to define all possible trajectories and all combinations of them (i. e. all permutations) and to find at least 100 examples to each of them for the training. Also the movement people could be very chaotic, when they waiting for a train on the platform. Such scenes have a lot of ambiguous scenarios. This implies construction of many very similar models, difficult classification and almost impossible detection of abnormal behaviour.

In an opposite case, we select a simple scene. For example, scene with an escalator is very well defined. In such scenes, objects usually moves with the same speed and in the same direction. However these trajectories are simple, they provide enough information for detection of injury or some other security threats.

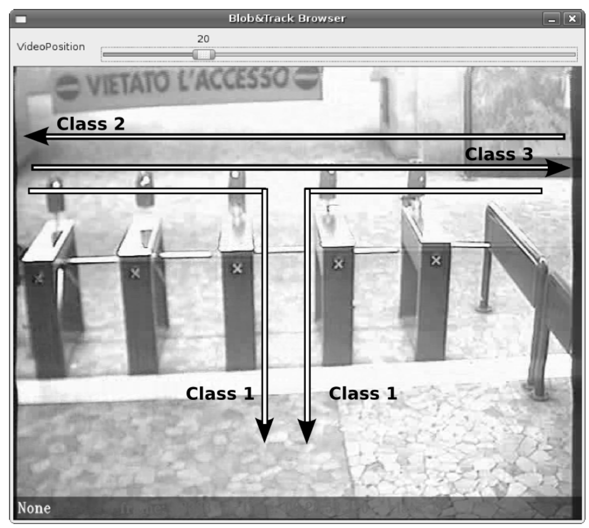

Figure 3: *Definition of normal trajectories in scene.*

For experiments was chosen the camera 3 from an underground station in Roma provided by [car 2008]. The scene is shown in Figure 3. The definition of normal trajectories was painted into picture for illustration.

5. EVALUATION

Two experiments were performed with recordings from an underground station in Roma. As a reference tracking algorithm for both of experiments the OpenCV [Intel 2007] tracker was used.

The first one was with filtered trajectories, so the scene contained only well defined trajectories.

There were defined three common trajectories. First, persons arrive from left or right of the scene, pass through turnpikes and leave the scene. Second, a person arrives from left and passes to the right and leaves the scene. The third was defined as opposite to the second normal trajectory, see Figure 3.

Classification was performed with about 400 trajectories in the training set and about 400 trajectories in the evaluation set with accuracy 91.92 %. The criteria for selecting corresponding class was a maximal match with a model. More detailed results with resolution of a True Positive, False Positive, True Negative and False Negative Classification are shown in following table.

class	TP [%]	TN [%]	FP [%]	FN [%]
1	62.12	32.83	0.51	4.55
2	5.30	87.37	6.31	1.01
3	24.49	72.22	0.76	2.53

Table 1: *Results of classification on filtered data.*

Results of each model were thresholded by several threshold and from resulting data ROCs[1] were generated. ROC of model of class 2 is shown in Figure 4.

Threshold for each model is determined from its ROC. When trajectory evaluated by a model reaches the threshold it is considered to be normal. Abnormal trajectories, therefore, are those which response do not reach a threshold of any model.

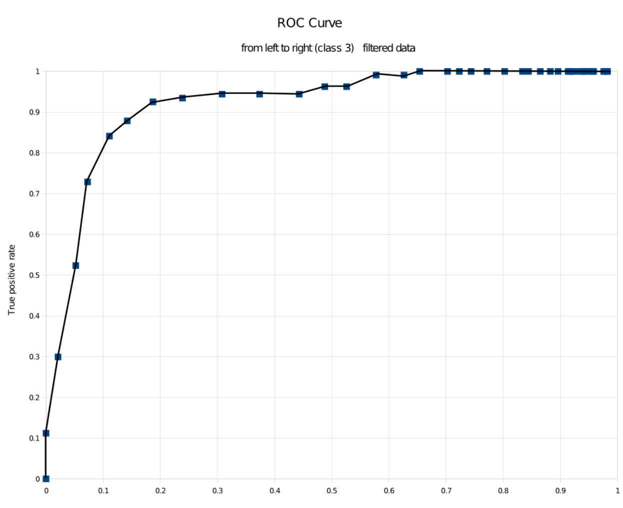

Figure 4: *ROC Curve for dark blue trajectory on filtered data.*

[1]Receiver Operating Characteristics

The second experiment was performed with trajectories, that were not preprocessed. The data set was containing also "insane" trajectories which was not corresponding to any person or to any other object in the scene. The noisy trajectories, we were unable to annotate, were modelled as class 4.

Using the training set corresponding to the first experiment and the evaluation set with about 10 000 trajectories, the accuracy was then only 37.63%. More detailed results are shown in the table below.

class	TP [%]	TN [%]	FP [%]	FN [%]
1	1.91	74.54	22.97	0.59
2	0.20	91.02	8.74	0.04
3	0.94	90.71	8.19	0.15
4	34.37	4.06	0.10	61.47

Table 2: *Results on non-filtered data.*

The ROC curve for class 3, e.g. corresponding to first experiment, is shown in Figure 5.

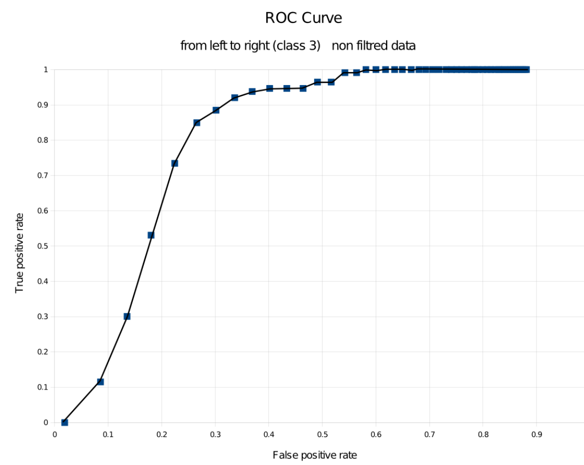

Figure 5: *ROC Curve for dark blue trajectory on non filtered data.*

6. APPLICATION

The attached screen shot in Figure 6 shows an example of the classification.

In the figure, there is a lady going through turnpikes. The trajectory complexity grows very fast with the complexity of the scene as it is noticeable in the first half of lady's trajectory (in the right half of the figure). The lady probably tried to use ticket on the first turnpike and after machine refused the ticket, she tried to use next turnpike.

In opposite part of trajectory (in the bottom part of the figure) is noticeable problems with tracking algorithm [Intel 2007], which systematically gives incorrect trajectory at its end.

Figure 6: *Example of trajectory classification*

7. CONCLUSIONS

In this paper, a method for trajectory classification has been presented. Surveillance camera recordings are processed to obtain objects trajectories. Some of the trajectories are annotated to create Hidden Markov Models. The method can provide information about the degree of correspondence of trajectory to each model. Performed experiments approved that the classification might be seriously successful when following conditions are met.

The necessary prerequisites for correct classification of a normal and abnormal behaviour are well defined (training) trajectories (e.g. defined in a simple scene or sub scene), well annotated abnormalities in trajectories and accurate object tracking technique. In this way, the proposed method enables also detection of objects with unusual behaviour.

8. ACKNOWLEDGEMENT

This work was supported by European project Caretaker (FP6-027231) [car 2008].

9. REFERENCES

AGGARWAL, J. K., AND CAI, Q. 1999. Human motion analysis: a review. *Comput. Vis. Image Underst. 73*, 3, 428–440.

BISHOP, C. M. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

2008. Caretaker. Information Society Technologies. http://www.ist-caretaker.org/.

CHAN, A. B., AND VASCONCELOS, N. 2005. Mixtures of dynamic textures. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, IEEE Computer Society, Washington, DC, USA, 641–647.

CHELLAPPA, R., ROY-CHOWDHURY, A. K., AND ZHOU, S. K. 2005. *Recognition of Humans and Their Activities Using Video*. Morgan Claypool, USA.

INTEL. 2007. *OpenCV (Open Source Computer Vision) Library*.

OHYA, J., UTSUMI, A., AND YAMATO, J. 2005. *Analyzing Video Sequences of Multiple Humans Tracking, Posture Estimation and Behavior Recognition*. Kluwer Academic Publishers, Norwell, MA, USA.

ROWEIS, S., AND GHAHRAMANI, Z. 1997. A unifying review of linear Gaussian models. Tech. rep., 6 King's College Road, Toronto M5S 3H5, Canada.

STARNER, T. 1995. *Visual recognition of American Sign Language using hidden Markov models*. Master's thesis.

STAUFFER, C., AND GRIMSON, W. E. L. 2000. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*, 8, 747–757.

YOUNG, S., ODELL, J., OLLASON, D., WOODLAND, P., AND JANSEN, J. 2006. *The HTK Book*.

ČERNOCKÝ, J. 2003. Hidden markov models – an introduction. Tech. rep., Brno University of Technology, Faculty of Information Technology.

ABOUT THE AUTHOR

Jozef Mlích is a last year graduate student at Brno University of Technology, Department of Computer Graphics and Multimedia. His contact email is xmlich02@stud.fit.vutbr.cz .

Petr Chmelař is a Ph.D. student at Brno University of Technology, Department of Information Systems. His research aims to the knowledge discovery from the multimedia data. His contact email is chmelarp@fit.vutbr.cz .