Trajectory Classification

AMIDA technology package description

Jozef Mlích, Jiří Ševcovic

Graph@FIT
Brno University of Technology
Faculty of Information Technology
Božetěchova 2
612 66 Brno, Czech Republic

Developers:
Jozef Mlích, Pavel Zemčík, Petr Chmelař,
Vítězslav Beran, Leoš Jiřík, Ondřej Vrana

Zemčík Pavel, zemcik@fit.vutbr.cz, responsible for AMIDA project
Herout Adam, herout@fit.vutbr.cz, responsible for Graph@FIT group

Last modified: December 23, 2009

Contents

1 Motivation

In AMIDA project was created technology for classification data from video source or a hardware device (Wiimote) to pre-annotated classes. Using this technology was tested on several applications - detection of dangerous conduct of person in the metro station, recognizing the gestures of people in video, voice activity detection based on gestures of the speaker and recognition of gestures using equipment Wiimote. Users can on the basis of sufficient quality data source, divided into required classes, create recognizer. The recognizer divides new data into marked classes, or earmarks the abnormal class that his behavior does not meet any indication class. Practically it is possible to use this technology in security systems, or any other applications where is necessary to divide the data source into predefined classes of behavior. In this technological package is available the aplication for training and recognizing, annotation tool and support for the Wiimote device, including user interface.

The trajectories in general describes spatiotemporal activity of objects. The main assumption is, that trajectory itself contain certain interesting information about object behavior. In case of surveillance applications, the observed objects are mostly humans and luggage in mass transit. In our work, we select simple scenes, in which is the behavior of people is easy predictable. In such scene we define trajectory classes, each of them denotes one standard situation. For example the class of people going trhu turn pikes and class of other people. The HMMs are trained on such classes. After that is determined the threshold for each model. The trajectories which not fits any model are considered to be abnormal. By normalisation and transformation to 3D could be classifier generalized for any scene containing the same scenario. It is also possible to use velocity and acceleration as features, which denotes better abnormal events that we want to spot.

An example of such scene is shown in Figure 1. The object passes thru scene are defined as follows. Object enter the scene and passes thru turnpikes. Object enter the scene at left side and leave scene at right without passing the turnpikes. Object enter the scene at right and leave scene at left without passing turn pikes.

Another application of trajectory analysis is related to automatic meeting data processing. The information retrival systems are using combination of sources to obtain result e.g. events spotted in trajectories. An example of information retrival system is speaker diarization, where the human activity

Figure 1: Surveillance scenario

is used in detection. We defined following classes: Speech supporting gestures (SSG) and non-SSG gestures. The anotation is obtained from unidirection microphone. The progress of gesture and detection of hand position is shown in Figure 2.

The last field of study is gesture recognition in user interfaces. The trajectory of gesture is obtained from accelerometer in wii remote controler device. For gesture recognition from trajectory was evaluate the accuracy of DTW [5] and HMMs [2]. For use in operating system was developed user interface, which allows various scenarios of human computer interaction. [5]

2 Features

The around described software requires standard desktop pc configuration with windows or linux.

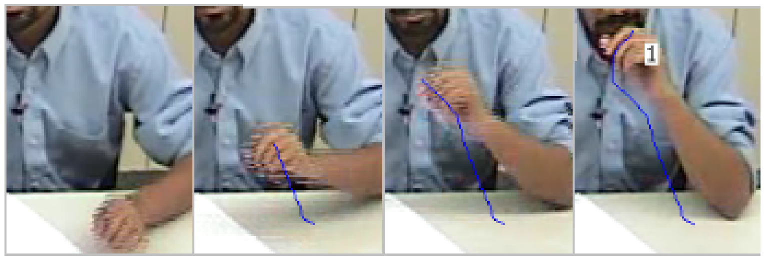

Figure 2: Example of non-SSG gesture

Figure 3: Wiimote based user interface

3 Credits

Contact persons: Pavel Zemčík, Adam Herout, Jozef Mlích.
Developers: Jozef Mlích, Pavel Zemčík, Petr Chmelař, Vítězslav Beran, Leoš Jiřík, Ondřej Vrana
Authors of this document: Jozef Mlích and Jiří Ševcovic.

4 Technical Description

In this part, we generaly describes methods, which are important and useful for classification of trajectories. First of them is DTW[1]. It is an algorithm for measuring similarity between two sequences which different length. For

[1] Dynamic Time Warping

instance, similarities in walking patterns would be detected, even if in one video the person was walking slowly and if in another he or she were walking more quickly, or even if there were accelerations and decelerations during the course of one observation.

The DTW could be described by Equation 1, where all samples m,n of sequences are compared by $d(m,n)$ and acumulated into array of buffers $g(prev)$, which defines the sum of best distances to refference sequence. The distances are modified by weight $w(k)$ of transition.

$$g(m,n) = \forall \max_{prev} [g(prev) + d(m,n)w(k)] \tag{1}$$

Next important term is GMM[2]. It is a probabilistic model for clustering and density estimation using a mixture distribution. According to set of example trajectories are created models which represent anotated classes. The correspondence of model and sample sequence is given by mixture of multivariate gaussian functions. The likelihood of mixture is defined by Equation 2, where a_i is weight for gaussian given gaussian function and $f_{Y_i}(X)$ is gaussian function defined by Equation 3.

$$f(x) = \sum_{i=0}^{n} a_i f_{Y_i}(X) \tag{2}$$

The k defines the dimension of feature vector x. The Σ and μ is covariance matrix and mean value of gaussian distribution.

$$f_Y(X) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)} \tag{3}$$

HMMs[3] approach belongs to supervised learning and statistical modelling methods for sequential data. Example of an HMM is shown in Figure 4. The sample model is described as a graph with four internal and two marginal states connected by (oriented) transitions. Moreover, there are six output vectors associated in Figure 4.

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).

[2]Gaussian Mixture Model
[3]Hidden Markov Models

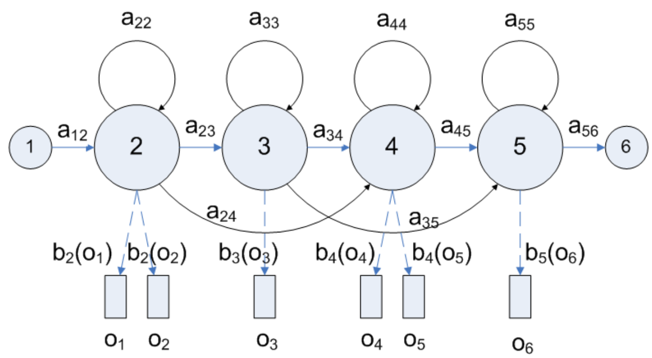

Figure 4: Example configuration of a Hidden Markov Model

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:

$$i^* = arg \max_i P\left(c_i|O\right) = arg \max_i \frac{P(O|c_i)P(c_i)}{P(O)} \qquad (4)$$

We use Bayes theorem in (4), 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 $b_j\left(o(t)\right)$, as illustrated in Figure 4.

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 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 5.

$$P\left(O,X|M\right)=a_{x(o)x(1)}\prod_{t=1}^{T}b_{x(t)}\left(o_{t}\right)a_{x(t)x(t+1)} \tag{5}$$

Viterbi algorithm (defined by Equation 6) 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.

$$P^{*}\left(O|M\right)=\max_{\{X\}}P\left(O,X|M\right) \tag{6}$$

For training the model M_i, corresponding to the trajectory class c_i, the Baum-Welch algorithm is used. It is a generalized expectation-maximization algorithm defined by Equation 7 that modifies weights of transitions and statistics of the models.

$$P\left(O|M\right)=\sum_{\{X\}}P\left(O,X|M\right) \tag{7}$$

4.1 Implementation

The software implementation of technology consist of following general usable packages. First is a set of unix shell scripts for HMM training based on HTK toolkit (http://htk.eng.cam.ac.uk/). Second the C++ library for online recognition of HMM in streamed feature vector.

Next, are included the applications for recognition of Speech Supporting Gestures, more concrete application for tracking of two persons in meeting scenario and gmm classifier used for recognition. (Visual Studio/C++)

The software related to surveillance scenario includes application browser and evaluation scripts. The browser allows to display trajectories in video sequences. The evaluation scripts are based on HMM unix script mentioned above.

The software related to Wiimote gesture recognition includes the scripts for evaluation of HMM and DTW accuracy, moreover application for connection of user interface control with gesture recognition module. (Windows, cygwin, mingw, unix)

4.2 Extendability

The wiimote user interface interaction scenario is easy extendable to any applications with DBUS (eventually DCOP) interface. User can define own gestures and it is possible to prepare set of standart gestures. It is also possible to extend functionality to recognition of combination of gestures. The fine tunning of user interface accroding to requirements of certain application.

The meeting scenario assumes gesture recogntion without device. It is possible to modify the meeting scenario for recognition of complex gestures or combination of static (e.g. symbols shown by hand) and dynamic gestures (e.g. shaking by hand, pawing). Moreover it seems to be natural connecting hand gesture recognition with user interface control. The problem for solving is propper division of gesture trajectory for compound gestures recognition.

The surveillance scenario suggest big ammout of objects interactions. The trajectory classification problem is easy to extendable to analysis of these interactions.

The accuracy of event recognition in all scenarios with video based data is strongly dependent on tracking. The state of the art tracking techniques could track the objects only in very limited conditions, therefore common sense suggest to improve tracking algorithms for purposes of trajectory analysis.

5 Technical parameters

- Processor: Intel Pentium 4, 2.8GHz, 2 GB RAM.
- Operating system: Windows or Linux
  (depends on scenario and application)
- Memory consumption: 30 MB
- CPU Usage: 0.1% to 40%
  (depends on algorithm and scenario)
- accuracy
  - DTW in wiimote scenario 70 %
  - GMMs in Speech supporting gestures scenario 62 %
  - HMMs in suveillance scenario 90 %

    - HMMs in wiimote scenario 85 %

6 Attachements

The package includes following attachments:

- Experimental software
- Demonstration videos
- Publications related to technology

The data (software, videos and papers) are organized according to relation with the publication describing the technology.

The technology describes follwing scenarios (and related technologies):

- Survaillance scenario (HMMs)
- Speech supporting gestures scenario (GMMs)
- Wiimote based user interface scenario (HMMs and DTW)

References

[1] Jozef Mlích. Trajectory classification. In *Proceedings of the 14th Conference STUDENT EEICT 2008*, Volume 2, pages 211–213. Faculty of Electrical Engineering and Communication BUT, 2008.

[2] Jozef Mlích. Wiimote gesture recognition. In *Proceedings of the 15th Conference and Competition STUDENT EEICT 2009 Volume 4*, pages 344–349. Faculty of Electrical Engineering and Communication BUT, 2009.

[3] Jozef Mlích and Petr Chmelař. Trajectory classification based on hidden markov models. In *Proceedings of 18th International Conference on Computer Graphics and Vision*, pages 101–105. Lomonosov Moscow State University, 2008.

[4] Jozef Mlích, Pavel Zemčík, and Leoš Jiřík. Trajectory classification using hmms. In *WSCG 2009 Communication Papers*, pages 67–72. University of West Bohemia in Pilsen, 2009.

[5] Vrana Ondřej. WiiMote as Input Device.Bachelor's thesis, Faculty of Information Technology, Brno University of Technology, 2009.