Human Activity Recognition in Archaeological Sites by Hidden Markov Models

Abstract

The problem of automatic recognition of human activities is among the most important and challenging open areas of research in Computer Vision.

This paper presents a methodology to automatically recognize the human activities embedded in video sequences acquired in outdoor environments by a single large view camera.

The recognition process is performed in two steps: first of all the body posture of segmented human blobs is estimated frame by frame and then, for each activity to be recognized, a temporal model of the detected postures is generated by Discrete Hidden Markov Models.

The system has been tested on image sequences acquired in a real archaeological site meanwhile actors perform both legal and illegal actions. Four kinds of activities have been automatically classified with high percentage of correct decisions. Time performance tests are very encouraging for using the proposed method in real time applications.

1. Introduction

Automatic recognition of human activities is one of the most important and interesting open area in computer vision. Unfortunately, a large portion of open literature in this field is devoted to activity recognition in limited know spaces where the human subject dominates the image frame so that the individual body components (head, hands, etc.) can be reliably detected. Detailed reviews of these works can be found in [6,7].

Few works dealt, instead, with the problem of human activity recognition in large areas. CMU's Video Surveillance and Monitoring (VSAM) project [1] and MIT AI Lab's Forest of Sensors Project [2] are two of the most appreciate examples of recent research efforts in this field.

In [2], the patterns and their activities are learned by motion analysis: some parameters as position and size of the patterns, and furthermore, velocity and direction of their motion allow the monitoring of humans and cars in a parking area. In [1], measurements based on a simple skeleton of the target are used to distinguish running people from walking ones.

Other considerable works in this area, like Pfinder [3] and W4 [4], try to classify humans and their activities by detecting features such as hands, feet and head, tracking and fitting them to an a prior human model.

In the VIGILANT project [5], an Hidden Markov Model approach has been employed to model the common event behaviours typical of a car-park environment. In this case velocity and width-to-height ratio of people and vehicle are supplied as input to an HMM procedure.

The analysis of the related works reveals that these algorithms for large area monitoring can recognize very simple activities like vehicle and person entering and exiting form a parking area, people running or walking and so on. The automatic recognition of these simple actions could not be adequate to meet the increasing requirements of surveillance applications such as the monitoring of parking areas, archeological sites or public buildings, etc. In these cases different algorithms able to discriminate between legal and illegal activities are required: for example the car theft cannot be prevented using only information about the car exiting from a parking area.

In this paper we propose a new approach to recognize illegal activities performed in a real archaeological site. A wide country area, where public or private plots of land are normally frequented by farmers, is continuously monitored with a large camera view. The problem is that, often, these lands are visited by burglars that steal the precious objects retained under the soil. The considered activities are very complex and the technical constraints due to large camera view impose some objective limitations to the applications of the well known related techniques. In this application context a surveillance system has to cope with a number of problems. First of all, the constraint of large camera view gives rise to images in which human beings cannot be easily segmented with all the body parts clearly visible. For this reason the classical methods of posture estimation based on skeleton measures cannot be applied. We propose an approach of posture estimation that works on binary

patches extracted from the images containing human blobs. The horizontal and vertical histograms of human blobs are computed and supplied as input to an unsupervised clustering algorithm. The Manhattan distance is used for both clusters building and run-time classification. The second problem is to understand if the sequences of distinguishable postures can be used to characterize different legal and illegal activities. In this work a statistical approach based on Discrete Hidden Markov Models is applied for building the models of four activities using a number of different example sequences of detected postures. The experimental results have demonstrated the effectiveness of the proposed approach to learn the four models and to generalize different executions of the same activities. The last point that has been addressed in this work concerns the ability of the method to recognize in a long test sequence the beginning of the known activities. We have used a sliding window that has been overlapped to the test sequence to extract a fixed length observation sequence provided to the behavior classification step. The proposed approach has been validated using 165 long test sequences acquired in a real archeological site.

In the rest of the paper, first a description of the generative model is presented and then the proposed activity recognition approach is detailed (section 2). Finally the experimental results obtained on real image sequences acquired in the archaeological site meanwhile actors perform both legal and illegal actions are reported (section 3).

2. Generative Model

HMM is a very useful technique for learning and matching activity patterns. It is characterized by a finite set of *states*, each of which is associated with a (generally multidimensional) probability distribution. Transitions among the states are governed by a set of probabilities called *transition probabilities.* In a particular state an outcome or *observation* can be generated, according to the associated *probability distribution*. It is only the outcome, not the state, visible to an external observer and therefore states are ``hidden'' to the outside; hence the name Hidden Markov Model. In order to define an HMM, the following elements are needed:

- The number of *states* of the model, *N*.
- The number of *observation* symbols in the alphabet, *M*.
- The set of state *transition probabilities* $A = \left\{ \begin{matrix} a_{ij} = P\{q_{t+1} = j \mid q_t = i\}, \\ 1 \le i, j \le N \end{matrix} \right\}$ where q_t denotes the current state.
- The set of *distribution probabilities* in each of the states $B = \left\{ \begin{matrix} b_j(k) = P\{o_t = \nu_k \mid q_t = j\} \\ 1 \le j \le N, \quad 1 \le k \le M \end{matrix} \right\}$ where ν_k denotes the k[th] observation symbol in the alphabet (CODEBOOK) and o_t the current parameter vector.
- The set of initial state distribution, $\pi = \{\pi_i\}$ where $\pi_i = P\{q_1 = i\}, \quad 1 \le i \le N$.

Therefore we can use the compact notation $\lambda = (A, B, \pi)$ to denote an HMM with discrete probability distributions. If the transition probabilities matrix A contains only non zero elements then the HMM is called Full Connected otherwise it is called Partially Connected. Full Connected HMMs can model more complex statistical processes but, at the same time, they are harder to manage because a large set of parameters has to be defined.

In figure 2 a full connected HMM with three states and a 2 dimensional codebook is represented.

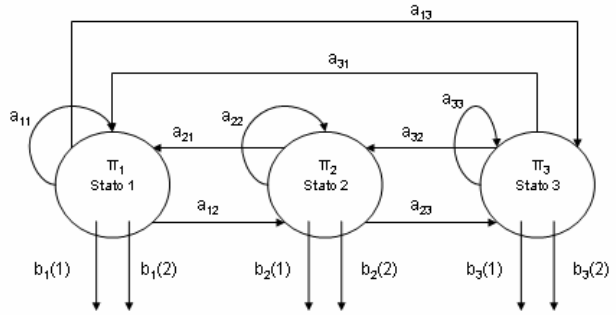


Figure 2: A full connected HMM with three states and two obserbvation symbols (N=3, M=2)

To recognize an observed symbol sequence $O = O_1, O_2, ..., O_T$ the probability of the observation sequence given the model is computed using Bayes' rule as $P(O \mid \lambda)$ and it is evaluated using the forward algorithm presented in [14]:

$P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i)$ where the forward variable $\alpha_t(i) = P(O_1, O_2, ..., O_t, s_t = q_i \mid \lambda)$.

To train an HMM to recognize the observation sequence O, the parameter set λ that maximizes $P(O \mid \lambda)$ must be estimated from the training data. The Baum-Welch algorithm [14] is used to iteratively obtain an estimate of λ that is guaranteed to locally maximize $P(O \mid \lambda)$. The Baum-Welch procedure defines three variable:

1)	$\beta_t(i) = P(O_{t+1},...,O_T, s_t = q_i \mid \lambda)$, (backward variable)
2)	$\gamma_t(i) = P(s_t = q_i \mid O_1,...,O_T, \lambda) = \frac{\alpha_t(i)\beta_t(i)}{P(O \mid \lambda)}$
3)	$\xi_t(i,j) = P(s_t = q_i, s_{t+1} = q_j \mid O_1,...O_T, \lambda) = = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{P(O \mid \lambda)}$

and, given an estimate of λ a better estimate λ' can be obtained as follow:

$$a'_{ij} = \frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)} \qquad b'_j(k) = \frac{\sum_{t=1,O_t=v_k}^{T-1}\gamma_t(i)}{\sum_{t=1}^{T-1}\gamma_t(i)}$$

$$\pi'_i = \gamma_1(i)$$

Learning converges to local maximum of $P(O \mid \lambda)$ when $\lambda = \lambda'$.

In the last years the Baum-Welch re-estimation formula has been extended to train an HMM with multiple sequences. The details can be found in [15].

3. Modeling Activities Using HMM

The generative model is applied to binary patches containing human blobs. In order to extract these patches a people segmentation algorithm is required. Since the description of this algorithm is beyond the scope of this paper, we give only a brief overview in the next subsection. We concentrate our attention on the details of the behavior classification using the generative model in the successive subsection.

3.1 People Segmentation

An algorithm for people segmentation has been implemented by applying a sequence of steps that are motion detection, shadow removing and object classification.

A modification of the approach proposed by T. Kanade, T. Collins and A. Lipton in [15] has been implemented for the motion detection task. In the training phase for each pixel a running average $\overline{p_n}$ and a form of standard deviation $\overline{\sigma_{p_n}}$ are maintained. In the test phase, a pixel is labeled as foreground if its intensity value differs from the running average two times more than the standard deviation. To cope with the problem of the sensitiveness to variations of the illumination conditions, a frequent update of the information about the running average and the standard deviation of all the background pixel values is carried out using an exponential filter as suggested in [15].
In order to obtain a significant shape of the extracted foreground objects a shadow removing algorithm has been applied. It is based on the assumption that shadows are half-transparent regions which retain the representation of the underlying background surface pattern. The implemented algorithm tries to detect moving regions with a texture substantially unchanged with respect to the corresponding background. The texture correlation has been evaluated by using the photometric gain between the model background image and the current image. Areas that present the same texture correlation are labeled as shadows and removed (see [19] for greater details).

The patches containing human blobs are extracted applying a sliding window on the foreground binary image that is provided to a neural classifier trained to detect human shape. In cases of multiple persons in the scene the classification algorithm provides different patches for each of them. Besides, a tracking procedure based on motion information (direction, velocity, position) has been implemented. In this way the following step will be able to track and recognize different behaviors that are currently in progress in the scene (see [20]).

3.2 Behavior Classification

The behavior classification algorithm executes two steps: first of all the human body postures have to be estimated in each frame and then the temporal sequence of detected postures has to be provided to the generative model described in section 2.

In the pose estimation step horizontal and vertical histograms of the binary shapes are evaluated and supplied as input to an unsupervised clustering algorithm named BCLS (Basic Competitive Learning Scheme) [18].

In the BCLS training phase, the number of clusters is a priori fixed by the user and the representative w_j of i-th class is randomly set. When a new training vector

x is presented, the algorithm identifies the winner cluster, i.e. the cluster with the higher proximity measure (lower distance), and all the representatives $\mathbf{w_j}$ are updated. The updating function of representatives suggested by [18] is

$$w_j(t) = \begin{cases} w_j(t-1) + \eta h(x, w_j(t-1)) & w_j \text{ is the winner} \\ w_j(t-1) + \eta^I h(x, w_j(t-1)) & \text{otherwise} \end{cases}$$

where η and η^I are two learning rate parameters. In this work we have decided to set $\eta^I = 0$ in order to update only the representative of the winner cluster. The training vectors are submitted many times and the algorithm is stopped when

$$\|\mathbf{W}(t) - \mathbf{W}(t-1)\| < \varepsilon \cdot$$

After the training phase, the algorithm is able to distinguish as many different postures in the training set, as the number of cluster required by the user.

In the BCLS testing phase an unknown vector **y** is assigned to the closer cluster in terms of lower proximity measure. In this work the proximity measure among two postures Im1 and Im2 is calculed as follow:

$$D(Im1, Im2) = d_1(X1,X2) + d_2(Y1,Y2)$$

where d_1 and d_2 are the Manhattan distances between the horizontal and vertical projections respectively.

The recognition of human behavior is then performed by statistical analysis of the temporal sequence of postures associated to a person in the scene. The number of different detected postures determines the number of the HMM codebook symbols (i.e the possible state values M). Each activity is associated to an HMM: this means that the number of HMM is always equal to the number of different activities of interest. The choice of the number of states N depends on the complexity of the process to be modeled: choosing an high number of states there is the risk of making the model too specific; on the other side with a small number of states there is the possibility of having indistinguishable activities. No general rule has been provided in literature for solving this problem. A proper tradeoff has to the done experimentally. Because of the complexity of the problem we deal with, we have selected fully-connected HMMs rather than partially connected ones. This choice has been made by observing that the fully-connected HMMs can model a wider range of process taking advantage from their more numerous parameters.

In the training phase the parameters of each HMM are updated in order to maximize the output probability of the training sequences. In this phase, the training procedure based on the multiple observation sequence proposed in [16] has been used. This training solution has been adopted considering that different people perform the same activity in different ways. The algorithm proposed in [16] expresses the multiple observation probability as a combination of individual observation probabilities. In particular we have implemented a generalizing Baum's auxiliary function and we have built an associated objective function using Lagrange multiplier method. For each different activity an HMM model λ_i ,as described in section 2, has been generated.

In the test phase unknown sequences are provided as input to the HMMs. The probability to have the activity A given the observation sequence X of postures is computed by evaluating the forward backward probability. A decision criterion based both on maximum likelihood measure

$$A^* = \arg\max P(X|\lambda_i)$$

and a set of proper thresholds to manage unknown behavior has been introduced. Indeed each HMM has associated a threshold equal to the minimum probability value obtained during the training phase.

The sequence X of posture observations is labeled as activity A if both its corresponding HMM gives the maximum likelihood measure among the whole set of HMMs and at the same time this probability value is greater than the relative HMM threshold. If this second condition is not satisfied the observation X cannot be associated to any of the known activities and is labeled as unknown.

The length of each observation sequence supplied to the HMMs is fixed in both training and testing phases and it has to be experimentally evaluated. Generally too short sequences are not enough to characterize any relevant activity; on the contrary too long sequences in the generative model do not allow the generalization of different executions of the same activity.

In the training phase the observation sequences are segmented by hand whereas in the testing phase a sliding window (of the same length of training sequences) is used to cover the whole acquisition sequence.

4. Experimental Results

The proposed human activity recognition approach has been tested on real sequences acquired in an archaeological site. The images were acquired with a static TV camera Dalsa CA-D6. In order to consider only significant frames for the activity recognition process we have sampled the acquisition sequence tacking two frames per second. The software was implemented by using Visual C++ on a Pentium III 1 Ghz and 128 Mb of RAM.

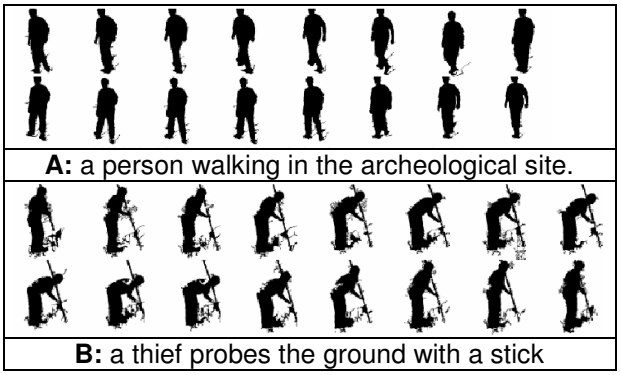
The archaeological site considered is a wide country area where some legal or illegal activities need to be discerned. In particular illegal activities are executed by people that first probe the subsoil using simple tools (such as sticks, tanks) and then they excavate to dig up some attracting objects.

The people segmentation algorithms produces for each person in the scene a binary patch of 175x75 pixels. Starting from these patches, the BCLS algorithm detects three kinds of different postures: "standing", "squatted" and "bent". One example of each detected posture can be found in figure 3. Sequences composed by a temporal succession of these three postures are supplied as input to the HMMs in order to identify 4 kinds of activities:

1. Walking
2. Probing the subsoil by a stick
3. Damping the ground with a tank
4. Picking-up some objects from the ground.

The first activity, the simpler one, is legal; while the remaining ones are more complicated and illegal. The figure 4 shows some frames for each of the possible sequences of the different activities. In particular the second and the third activities are very similar: they are composed by sequences of the same two postures, but with different temporal variations. The statistical modeling step is then composed by 4 HMMs. Each HMM is associated with a different kind of activity and it is trained with three different examples (performed by different people) of the associated activity.

Figure 3. Three fundamental postures classified in the archaeological site.



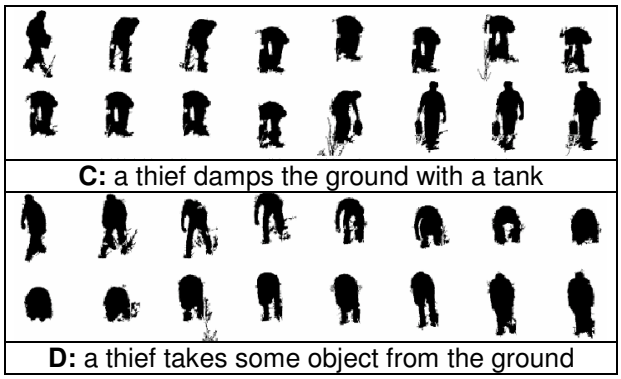


Figure 4. Some frames extracted from 4 of the 12 sequences (4 activities x 3 sequences) used to train the HMMs

The training set, composed by 4 x 3=12 sequences is not changed during all the experiments described below. Each training sequence consists of 50 frames (so 50 is also the length of the sliding window used in the test phase). The experimental tests have shown that a greater number of training sequences decreases the generalization ability of the HMM, as asserted in [16].

In the first experiment, the system has been tested using 160 sequences. Each sequence contains one of the 4 activities to be recognized (just 40 for each kind of activity), but the beginning and the ending frames are not known. The length of the input sequence ranges from 400 to 1500 frames. If N_{TOT} is the total number of frame in each test sequence and n is the length of the sliding window then $N_W = N_{TOT} - (n-1)$ is the number of windowed observation sequences O_w supplied as input to the HMMs for each test sequence.

An activity is recognized in a test sequence when at least one of its observation sequence O_w extracted by the sliding window, satisfies the recognition procedure described in the previous section (bayesan criterion + adaptive threshold).

Table 1: The activity recognition results when the number of HMM states changes.

HMM States	Activity			
	Walking People	Probing People	Damping People	Picking up People
2	***40/40 100%***	***40/40 100%***	0/40 0.0%	***40/40 100%***
3	40/40 100%	25/40 62.50%	29/40 72.5%	40/40 100%
4	40/40 100%	25/40 62.50%	30/40 75%	40/40 100%

5	40/40 100%	28/40 70%	26/40 65%	25/40 62.5%
6	40/40 100%	27/40 67.50%	28/40 70%	40/40 100%
7	40/40 100%	27/40 67.50%	26/40 65%	40/40 100%
10	40/40 100%	20/40 50%	***31/40*** ***77.50%***	40/40 100%
12	40/40 100%	22/40 55%	28/40 70%	40/40 100%

In table 1 HMMs with 2-3-4-5-6-7-10 and 12 states have been tested in order to determine the optimal number N of HMM states in our application domain.

Each test sequence is given as input to the four HMMs with the same number of hidden states. For this reason, the results of every row of the table have to be considered altogether. Figure 5 shows the variation of the mean percentage of correct classification when the number of HMMs states changes. On the X axis there is the number of Hmm states, whereas, on the Y axis there is the mean percentage.

Figure 5 sums up the best classification results obtained with 4 and 6 hidden states. In this case the percentage of right classification is 84.37%. HMM with a larger number of states have not been considered because the HMM's theory [14] suggests the use of a number of states much smaller than the number of symbols in each observation sequence (50 in our case).

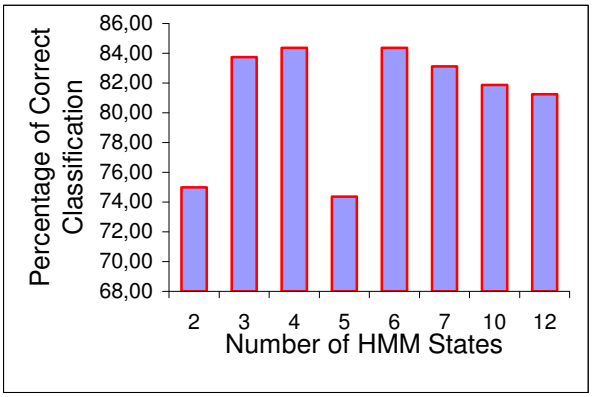


Figure 5: Variation of the mean percentage of correct classification when the number of HMMs states changes

In the second experiment, in order to further improve the classification results, the four HMMs with the best classification performances have been selected and tested on the same 160 sequences used in the experiment 1. Notice that in this case the performances of the proposed approach can change with respect to the ones reported in table 1, since both the relative maximum and corresponding threshold are used to classify each sequence.

For the sequences "walking people" and "Picking up People" two hidden states have been selected because, under the same conditions, a smaller number of states makes simpler the training and test algorithms. For the sequence "Probing people" two states have also been selected because this case is the only one that ensures a classification performance of 100%. For the sequence "Damping People" the HMM with ten states has been selected since it ensures the best classification performance (77.50%). The classification values relative to the selected HMMs are reported in cursive and bold type in table 1.

The mean percentages of correct recognitions of the experiment 2 are reported in table 2 whereas table 3 shows the relative scatter matrix. The results demonstrate the effectiveness of the proposed approach based on a combination of HMM with different state numbers. The scatter matrix shows that the system mistakes the activities "probing people" and "Damping People". Actually these two activities are very similar and hard to distinguish also for a human beings.

A further experiment was performed: we have supplied to the HMM architecture used in the experiment 2 a set of 5 sequences containing none of the 4 activities used in the training phase. In this case no false positives have been found (meaning that the threshold constraint relative to the winner HMM is never satisfied).

Finally, in order to evaluate the possibility of using the proposed approach for real time applications, some considerations about the computational load have been done. Each frame can be processed in about 14×10^{-2}s and the distribution of the computational load in the four subsystems is reported in table 4. In the first column the computation time for the whole people segmentation algorithm is reported, while we have distinguished the computational times for the posture estimation and the activity recognition in the second and in the third columns. The total amount allows the processing of 6 frames/sec. Of course, this is a satisfying result taking in account that normally the human movements are slow and that at the moment no specialized or parallel hardware has been considered.

Table 2: The activity recognition results when the best HMM architecture of the experiment 1 were used

Walking person HMM	Probing People HMM	Damping People HMM	Picking up People	Mean Percentage of correct

with 2 states	with 2 states	with 10 states	HMM with 2 states	classification
40/40 100%	29/40 72,50%	30/40 75%	40/40 100%	139/160 **86,87%**

Table 3: Details of the activity recognition results when the best HMM architecture of the exp. 1 was used

Scatter Matrix	HMM classification			
	Walking person	Probing People	Damping People	Picking up People
Walking person	**40**	0	0	0
Probing People	0	**29**	11	0
Damping People	0	10	**30**	0
Picking up People	0	0	0	**40**

Table 4: Distribution of the computational load

Segmentation	Pose Estim.	Activity recognition	**Estimated Total Time per frame**
~ $4x10^{-2}$s	~$1x10^{-1}$	~$5x10^{-5}$	**~ $14x10^{-2}$ s**

4. Conclusions and Future Works

In this paper we have presented a reliable approach to recognize complex human activities performed by human beings in wide outdoor environments. In particular we have addressed some of the problems concerning this kind of application domains.

Starting from the detection of moving people the proposed approach addresses the problem of recognizing four different activities from temporal variations of postures. The postures have been detected using an unsupervised clustering algorithm whose only demand is the number of expected cluster. The algorithm is able to separate the binary shapes in the required number of classes. Experimental results have demonstrate that the obtained classes correspond to the main postures assumed by people in our sequences.

Fixed length sequences (50 frames) of postures have been used both in training and test phase to model the four different activities and to classify new examples of the same behavior.

The experiments have demonstrated the effectiveness of using HMM as generative model to recognize activities based on sequence of temporal postures. Misdetections happen on the activities "Probing People" and "Damping People" that are very similar: they are characterized by the same postures with light variations in their temporal evolutions. Actually, it is difficult to distinguish between them also for human beings, unless the further recognition of objects carried by people is added to the system. This possibility has not been considered in this work since the same activity can be executed in different position with respect to the camera, and in many cases the objects could be occluded. The aim of this work was to demonstrate that a raw classification of postures and the analysis of their temporal evolution can be enough to recognize different legal and illegal activities.

Besides, the computational times have been evaluated for each step of the whole system: they are very encouraging for using the system in real time applications.

Future work will be addressed to evaluate how a larger number of postures can improve the results of the activity classification, also considering that the same position of a person can be perceived in a different way from the camera according to the relative orientations. Besides, we will face the problem of selecting variable length observation sequences from the test sequences, in order to overcome the constraint imposed in this work of having the same behavior in quite the same number of frames.

5. References

[1] R.T. Collins, A.J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, "A System for Video Surveillance and Monitoring", *Technical Report CMU-RI-TR-00-12*, Carnagie Mellon University, 2000.

[2] C. Stauffer and W.E.L. Grimson, "Learning Patterns of Activity Using Real-Time Tracking" *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 22, n.8, pp. 747-757, August 2000.

[3] C. Wren, A. Azarbayejani, T. Darrell, and Alex Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligenc*e, 19(7):780–785, 1997.

[4] Haritaolu, D. Harwood, and L. S. Davis, "W-4: Real-time surveillance of people and their activities," *IEEE Transactions Pattern Analysis, and Machine Intelligence*, vol. 22, no. 8, pp. 809-830, 2000.

[5] P. Remagnino and G.A. Jones, "Classifying Surveillance Events from Attributes and Behaviors" *in the Proceeding of the British Machine Vision Conference*, September 10-13, Manchester, pp. 685-694. ISBN/ISSN 1901725162 (2001)

[6] D. Ayers and M. Shah , "Monitoring human behavior from video taken in an office environment", *Image and Vision Computing*, Vol. 19 (12) (2001) pp. 833-846.

[7] M. Petkovic, W. Jonker and Z. Zivkovic, "Recognizing Strokes in Tennis Videos Using Hidden Markov Models", In proceedings of Intl. Conf. on Visualization, Imaging and Image Processing, Marbella, Spain, 2001.

[8] A. Galata, N. Johnson and D. Hogg, "Learning Variable Length Markov Models of Behaviour**",** *Computer Vision and Image Understanding : CVIU,* vol. 81, n. 3, pp. 398-413, 2001.

[9] Y. Wu and T. S. Huang, "Vision-Based Gesture Recognition: A Review", *Lecture Notes in Computer Science,* vol.1739, pp.103-115", 1999.

[10] A. D. Wilson and A.F. Bobick, "Parametric Hidden Markov Models for Gesture Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 21, n. 9, pp. 884-900, 1999.

[11] I. Cohen, A. Garg and T. Huang, "Emotion recognition from facial expressions using multilevel HMM", *Neural Information Processing Systems*, 2000

[12] M. Brand, N. Oliver and A. Pentland, "Coupled Hidden Markov Models for Complex Action Recognition", *In Proceedings of IEEE CVPR97*, 1996

[13] H. Ren and G. Xu "Human Action Recognition in a Smart Classroom" *in the Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FGR™02).*

[14] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Processing", *Proceedings of the IEEE,* vol. 77, pp. 257-286, 1989**.**

[15**]** T. Kanade, T. Collins and A. Lipton, "Advances in cooperative multi sensor video surveillance, *DARPA Image Understanding Workshop*, Morgan Kaufmann, Nov. 1998, pp.3-24

[16] X. Li, M. Parizeau, R. Plamondon, "Training Hidden Markov Models with Multiple Observations- A Combinatory Method", *IEEE Transactions on PAMI*, vol. PAMI-22, no. 4, pp. 371-377, April 2000

[17] C. Gurrapu and V. Chandran "Gesture Classification Using a GMM Front End and Hidden Markov Models", *in the Proceedings of the 3rd IASTED International Conference Visualization, Imaging and Image Processing*, September 8-10, 2003, Benalmadena, Spain pp.609-612.

[18] S. Theodoridis,., K. Koutroumbas,, "Pattern Recognition", *Academic Press*, *San Diego, 1999, ISBN 0-12-686140-4.*

[19] A.Branca, G.Attolico, A.Distante "Cast Shadow Removing in Foreground Segmentation". *In Proc. Int. Conf. on Pattern Recognition, 2002.*

[20] M. Leo, G. Attolico, A. Branca, A. Distante "People detection in dynamic images" *In the proceedings of the IEEE World Congress on Computational Intelligence (WCCI 2002), Honolulu, Hawaii, May 12-17, 2002*