Turk J Elec Engin, VOL.13, NO.3 2005, © TÜBİTAK

Control Charts Approach for Scenario Recognition in Video Sequences

Ersin ELBASI, Long ZUO, Kishan MEHROTRA,
Chilukuri MOHAN, Pramod VARSHNEY
Department of Electrical Engineering and Computer Science Syracuse University
Syracuse, NY 13244-4100
{*eelbasi, lzuo, kishan, mohan, varshney*}*@ecs.syr.edu*

Abstract

A new approach, based on control charts, is presented for the task of recognition of events and scenarios in video image sequences. For each image in the sequence, low level image processing and feature extraction steps result in feature descriptors for objects of interest detected in the images. Control charts analysis is then explored to classify the nature of the activity depicted by the temporal changes in these features over the image sequence. Scenario recognition with higher accuracy is achieved using this simple approach.

1. Introduction

In recent years, safety and security concerns have lead to a considerable increase in video surveillance and monitoring research efforts. These have found many military and civilian applications such as monitoring of banks, parking areas, buildings, department stores, and national borders. Most of the current monitoring systems use multiple cameras and human operator to detect unexpected scenarios, because in the realistic application it is difficult to monitor a large target area at once, and difficult to track moving object over a long time period. There exists demand for automatic scenario recognition. However, automatic video understanding is a difficult task and is an active area of research interest.

There are two main steps in scenario recognition:

1. Low-level processing, including background model generation, moving object detection, illumination removal, object tracking and background/template updating, resulting in the generation of a vector of features that abstracts the scenario; and
2. High-level processing, which uses the features to perform event classification and scenario recognition; this is the focus of the research presented in this paper.

Several methods have been used to process the features for scenario recognition. In particular, Bayesian approaches and Hidden Markov Models (HMM) have been extensively used to detect simple and complex events that occur in the scenarios. This paper shows that an alternative and simpler approach, based on *control charts,* is equally effective in detecting activities that occur in a scene.

This paper is organized as follows. A brief introduction to related works is presented in section 2. Section 3 describes the low-level processing steps used in our work. In Section 4 we present the high-level processing steps including control charts and other methods of scenario identification. Experimental results are represented in section 5, and conclusions in Section 6.

2. Related Works

Considerable research has been performed on scenario recognition, including many approaches that can be distinguished in terms of:

- model-based versus non-model-based;
- 2D versus 3D;
- single camera versus multi-camera;
- static camera versus dynamic camera;
- activities involving one person versus multiple persons; and
- offline versus online

detection. In recent years, several models based on finite state machines have been widely used in speech recognition, natural language processing, sign language and scene analysis. Using one camera, Hongeng *et al* [1, 2, 3] has recognized a sequence of several scenarios, called a multi-state scenario, using Bayesian network and Hidden Markov Models (HMMs). Hamid [4] tracks objects with color and shape based on particle filters, to extract features, and applies Dynamic Bayesian Networks to recognize events. Yamato [5] describes a new human behavior recognition algorithm based on HMMs; a sequence of frames is converted to a feature vector, and converted again to a symbol sequence by vector quantization. HMMs are trained, and the model that matches the event best is selected. Parameterized HMMs and coupled HMMs have been used to detect complex events such as the interaction between two moving objects.

For robust detection, successful feature extraction is essential. The most useful features in human tracking detection are found to be the width, height, color histogram and velocity. Chowdhury [6] experimented with detection of normal and abnormal events at an airport, and proposed a method to represent the activity of a dynamic configuration of objects by the shape formed by the trajectories of these objects. Object activities are represented as points on the 2D ground plane. Amer [7] has worked on detection of events such as walking, sitting, and standing, using simple object features such as the width and height to detect events. The difficulty in the above mentioned works is in differentiating between real moving objects and clutter such as trees blowing in the wind and moving shadows. Davis [8] has worked on reliable recognition of basic activities from the smallest number of video frames. He used probabilistic methods to detect simple activities such as walking, running and standing. Many other papers [7], [8], [9] also address simple scenario recognition based on probabilistic methods.

Our approach is different from the above methods currently in use. We use a rule-based system to categorize detected human activity into various classes, where the rules are obtained by control chart analysis. In essence, we treat the problem as analogous to controlling a manufacturing process. A process in control is analogous to the continuation of a sub-scenario and the time when the process goes out of control indicates that the tracked object is transitioning from one sub-scenario to another.

3. Low-level Processing

A system for robustly tracking objects is used for the indoor surveillance application. Background subtraction has been implemented to detect foreground objects associated with regions that exhibit small changes over time. We adopt a luminance contrast method [10], [11] to reduce the side-effect of background subtraction, for two reasons:

1. It saves computational effort by using just one channel in color images.

2. It removes much of the noise caused by luminance variations.

The original RGB space images from video camera are transformed to YUV coordinates. Null luminance values result in infinite contrast values that have no physical meaning, hence such values are changed to 1s. Values near zero are expected for background pixels, with larger values for brighter pixels.

A standard background subtraction method is used to extract the silhouette of the moving object. To implement this, the mean and the standard deviation of each pixel are computed in a series of images without any person. Then, a pixel is considered as belonging to a moving object if the difference between the mean and the current value of that pixel is higher than a certain threshold that depends on the standard deviation. The background is updated for each new image by recalculating the mean and standard deviation at all pixels.

In the work reported in this paper, we utilized a successive refinement strategy for object tracking. First, we applied a motion detection algorithm to locate the region in space containing a moving object. Then, the dynamic template representing changing image frames is matched in search regions to further localize object position accurately. We used an infinite impulse response (IIR) filter to update the template [13]. Once a best correlation match at the search block in current frame is found, it is merged with a previous template through an IIR filter to produce a new template for tracking in subsequent frames.

4. High Level Processing

A *Control Chart* is a tool for process monitoring, providing a popular method for quality control in manufacturing engineering. Each control chart indicates the variation in the values of some feature over time, with graphical depiction of the *upper* and *lower* control limits for that feature [14], [15]. There are two main parts in the high level detection by control charts: the identification of an activity or sub-scenario, and the recognition of when (in the sequence of images) the activity begins and ends.

Our simulations were restricted to scenarios in which the only objects of interest are human figures, although this restriction can be relaxed to accommodate more general scenarios. Several features are computed for objects detected in each image frame, including the height and width of the bounding box containing the object, and measures describing the relative distribution of foreground pixels in the upper, middle, and lower regions of the bounding box.

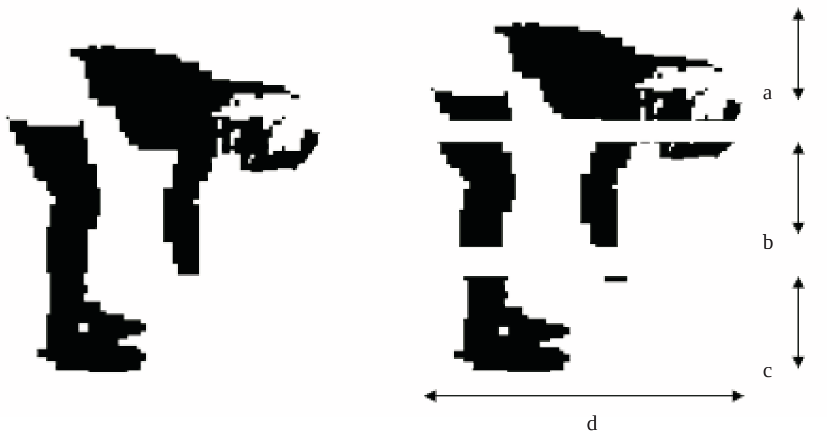
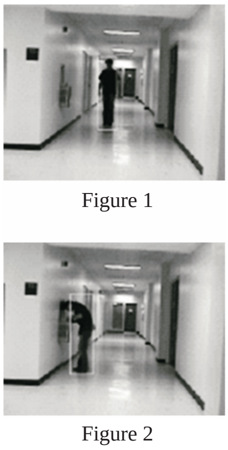


Figure1. RUD= a/ (a+b+c) , RMD= b/ (a+b+c) and RLD=c/(a+b+c).

The speed of a moving human is used to classify it as running or walking. Discrimination between three other classes (viz., standing, bending, and sitting) is performed using control charts constructed for four features, described below, of which the first is well-known and the other three are new features developed in this work that are easy to compute and helpful in the classification task:

- *Aspect ratio (AR)* = (width of the bounding box)/(height of the bounding box)
- *Relative Upper Density (RUD)* = The fraction of foreground pixels in the upper 33% of the bounding box
- *Relative Middle Density (RMD)* = The fraction of foreground pixels in the middle 34% of the bounding box
- *Relative Lower Density (RLD)* = The fraction of foreground pixels in the lower 33% of the bounding box



Feature	Walking	Bending
AR	0.2	0.3
RUD	0.3	0.2
RMD	0.4	0.5
RLD	0.3	0.3

Figure 2. Typical values of selected features for two postures: walking and bending.

The features used do not satisfy assumptions essential to use multivariate control charts, and classification accuracy obtained using a single multivariate control chart was found to be poor. Instead, more successful results were obtained on building four control charts, one for each feature.

Each control chart is constructed using the following procedure for each feature f_j. The mean and standard deviation of f_j values are first computed for all three classes to be determined, using the available training data. Then, for each class c_i, the upper and lower bounds associated with the control charts are obtained, using the equations

$$\textit{upperBound}\ (f_j\ , c_i) = \text{mean}\ (f_j\ , c_i) + \alpha\ (f_j\ , c_i\ )\ \times \text{standard deviation}\ (f_j\ , c_i)$$

and

$$\textit{lowerBound}\ (f_j\ , c_i) = \text{mean}\ (f_j\ , c_i)\ \text{-}\ \alpha\ (f_j\ , c_i\ )\ \times \text{standard deviation}\ (f_j\ , c_i)$$

Approximate values for each coefficient $\alpha\ (f_j, c_i)$ are found using an iterative improvement approach.

Each feature's control chart thus suggests that an object be placed into one of the classes; since we use four features, this results in 4-dimensional vector whose components indicate the class in which an object is placed according to each control chart. For example, the tuple [Standing, standing, Standing, Bending] indicates that the first three features place the object into the "Standing" class whereas the last feature's control chart places it in the "Bending" class.

Final classification uses the majority rule applied to such a vector, so that the above example will be placed in the standing class.

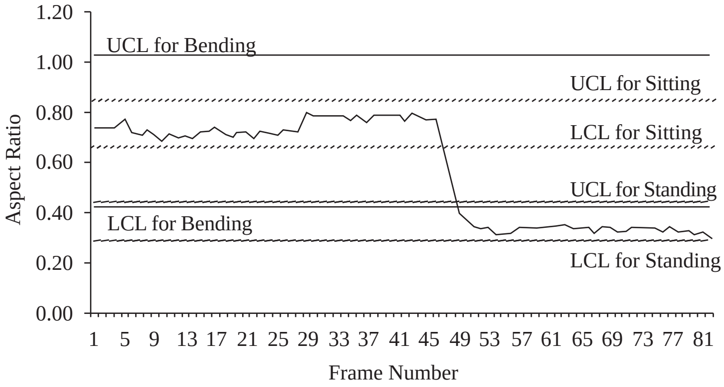


Figure 3. Control Chart for Aspect Ratio, showing upper and lower control limits (UCL and LCL) for three classes of activity.

Ties are broken by giving extra importance to an appropriate feature, determined as follows, and comparing the distance of the observation from the mean of the chosen feature.

A tie between 'Standing' and 'Bending' is broken in favor of 'Standing' if the value of RUD feature for the candidate object is closer to *mean*(RUD, Standing) than to *mean*(RUD, Bending).

A tie between 'Standing' and 'Sitting' is similarly broken using values for the AR feature.

A tie between 'Sitting' and 'Bending' is broken similarly using the RLD feature.

4.1. Detecting change of sub-scenario

The above classification algorithm, in which objects are placed into one of the classes based on feature values, is subject to potential errors due to image quality. Fortunately, many such errors can be corrected based on a temporal continuity assumption: it is highly unlikely that an object's activity classification fluctuates rapidly within successive time instants. Hence we detect a change of sub-scenario using a criterion based on a parameter $c > 0$ describing the minimum expected duration for any sub-scenario. If c consecutive decisions at times $t, (t-1), \ldots, (t-c+1)$ are all different from the decision being made at time $(t-c)$, then we conclude that a new sub-scenario had commenced at time $(t-c+1)$. Otherwise, we attribute the differences to noise and image quality, and presume that the sub-scenario has not changed.

5. Experimental Results

In our experiments we applied the above method to four video segments of different sizes. In each video, the subjects carried out various sequences of activities that included standing, sitting, and bending. These videos were combined to produce a total of 2550 frames. From a total of 2550 feature vectors, obtained from these frames, 1072 were used in the training and the rest of the 1478 vectors were used in testing the classification performance.

The classification accuracy of the new algorithm was 95.7% correct, i.e., 95.7% of the test cases were accurately classified. The performance of the control chart method, for determining the end of one sub-scenario and the beginning of another (as described in the previous section), is presented in Table 1.

Table 1. Sub-scenario recognition accuracy of Control chart approach.

Video	Number of frames	Number of sub-scenarios	Number of recognized scenarios
1	823	11	10
2	512	6	6
3	701	12	12
4	514	9	10

The table clearly shows that the accuracy of correct classification of sequence of sub-scenarios is very high and the control chart method results in excellent classification performance. Comparison with three other popular classification techniques on the same data also support the control chart approach: a backpropagation neural network resulted in 91.34 %, C5 algorithm yielded 92.86%, and naïve Bayes classifier 89.61%, whereas the control chart method gave 95.7% accuracy.

6. Conclusions

We have presented a new approach for monitoring activities of objects over time in video sequences, addressing the main goal of surveillance systems. The main feature of our approach is the use of control charts, applied to multiple features obtained by processing image data. Our system correctly detects activities and transition between them. These are very promising results and indicate that there is much to be said in favor of simple methods even when the problem is complex. Although this approach does not solve the problems arising in the scenario analysis context, we suggest that simple classification rules such as those based on control charts, should be integrated with other methods such as Hidden Markov Models, possibly using a successive refinement strategy.

Acknowledgement

This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD10-00-1-0352.

References

[1] S. Hongeng and R. Nevatia,“Multi-agent event recognition,” *IEEE International Conference on Computer Vision,* July 2001.

[2] S. Hongeng, F. Bremond and R. Nevatia, “Representation and optimal recognition of human activities,” *IEEE Computer Vision and Pattern Recognition*, June 2000.

[3] S. Hongeng, F. Bremond and R. Nevatia, “Bayesian framework for video surveillance application,” *International Conference on Pattern Recognition,* September 2000.

[4] M. Hamid, “ARGMode – Activity recognition using graphical models,” *Computer Vision and Pattern Recognition*, June 2003.

[5] J. Yamato, S. Kurakakae, A. Tomono, and K. Ishii, “Human action recognition using HMM with category separated vector quantization,” *Transaction of Institute of Electronics, Information, and Communication Engineers*, 1994.

[6] A. R. Chowdhury and R. Chellappa, “A factorization approach for activity recognition,” *Computer Vision and Pattern Recognition*, June 2003.

[7] A. Amer, “A computational framework for simultaneous real-time high-level video representation,” *Multisensor Surveillance Systems*, pp 149-182, July 2003.

[8] J. W. Davis and A. Tyagi, “A reliable inference framework for recognition of human actions,” *IEEE Conference on Advanced Video and Signal Based Surveillance,* July 2003.

[9] T. Moeslund and E. Granum, “A survey of computer-based human motion capture,” *Computer Vision and Image Understanding,* vol 81, March 2001.

[10] R. T. Collins, A. J. Lipton and T. Kanade, “A system for video surveillance and monitoring” *Proc. American Nuclear Society* (*ANS*), Eighth International Topical Meeting Robotic and Remote Systems, 1999.

[11] L. M. Fuentes and S. A. Velastin, “People tracking in surveillance applications,*” Proceedings 2nd IEEE International Workshop on PETS*, Kauai, Hawaii, USA, December 9, 2001.

[12] J. R. Quinlan, “C5.0: An informal tutorial,” *Rilequest Research*, http://www.rulequest.com/see5-unix.html, 2002.

[13] A. Lipton, H. Fujiyoshi, and R. Patil, “Moving target detection and classification from real-time video,*” IEEE Workshop on applications of Computer Vision*, October 1998.

[14] D. C. Montgomery, “Introduction to Statistical Quality Control’, John Wiley & Sons; 3rd edition (August, 1996).

[15] D. T. Pham and A. B. Chan, “Unsupervised neural networks for control chart pattern recognition,” *CIRP International seminar on Intelligent computation in Manufacturing Engineering*, 1998.