FEATURE SIMILARITY EVALUATION AND EVENT CLUSTERING FOR VISION SYSTEMS

W.S.K. Fernando[1], P.H. Perera[2], H.M.S.P.B. Herath[3],
M.P.B. Ekanayake[4], G.M.R.I. Godaliyadda[5], and J.V. Wijayakulasooriya[6]

Dept. of Electrical and Electronic Engineering, University of Peradeniya, Sri Lanka.
Email: {[1]sarith, [2]pramuditha, [3]samitha, [4]mpb.ekanayake, [5]roshangodd, [6]jan}@ee.pdn.ac.lk

ABSTRACT

An approach to perform video event classification by taking into account feature similarity through spectral clustering is proposed in the paper. Spectral clustering is used to identify different types of events and to define respective event signatures. Number of modifications to standard spectral clustering algorithm is proposed to make it suited for an event classification application. Dynamic Time Warping (DTW) is used to overcome the challenge of evaluating pairwise inter-event disparity values which is required for spectral clustering. Methods to determine initial parameters for spectral clustering are also proposed. Observed events are compared against event signatures to determine their event identity and to detect anomaly events. Event classification results obtained for a series of videos that involve human motion are discussed in the paper.

***Key words*:** Event classification, Signal comparison, Visual surveillance, Spectral clustering

1. INTRODUCTION

Video event classification and abnormal event detection are highly employed in modern autonomous vision systems. They are extensively used in real time applications such as in surveillance systems, traffic monitoring systems and in behavior analytic systems. Video event classification involves two main stages; object feature space construction and event classification through clustering. Proposed here is a comprehensive mechanism designed to cater the latter stage of event classification process for a real time application.

Extracted object features through foreground estimation and tracking [1] are compared based on their similarity in shape, and a disparity figure is evaluated. Inter-event disparity is used to identify possible event types in the scene using spectral clustering and event signatures are produced to represent them. Observed new events are compared against existing event signatures to determine the event identity. Events that do not match any of the event signatures are classified to be anomalous.

Even though problem of adopting spectral clustering for event detection is addressed in [2, 3], a robust method to determine initial parameters of spectral clustering, namely number of clusters and kernel variable σ, is not presented. Moreover, [3] proposes applying spectral clustering iteratively at each time instant which requires exhaustive computational power from processers' perspective thereby making it unsuitable for practical implementation. When evaluating inter-event disparity, which is a requirement of spectral clustering, Hidden Markov Models are used in [3]. However, due to partial matching of models, this approach results in a higher degree of fault identifications in the evaluation process. Further, Hausdorff distance proposed in [4] fails to capture the time non-linearities of trajectories making it unsuited to compare trajectory shapes when an object demonstrates temporal discontinuities in its trajectory.

Major contributions of the paper are as follows. This paper proposes spectral clustering for initial event signature generation phase. Dynamic Time Warping is introduced to evaluate inter-event distance to reflect event disparities irrespective of nonlinearities in time domain. An automated method to select Context dependent initial parameters for spectral clustering is also proposed. A statistical method is proposed to identify event identities by comparing event signatures with the observed event. Method for detecting anomalies too is described.

This paper includes 6 main sub sections. Section 2 includes an overview of the proposed system. Event disparity evaluation model is discussed in sections 3. Sections 4 and 5 elaborate the event classification and anomaly detection respectively. Results of the presented method are discussed in section 6 and paper concludes with section 7.

2. OVERVIEW

The methodology comprises of two phases; training phase and real-time phase. This process is illustrated in Figure 1. In the training phase, a training video consisting of different events is first subjected to foreground segmentation and tracking to generate feature traces that describe events. Methodology presented in [1] is used for this purpose. Features of each event are then subjected to spectral clustering to identify different possible types of events. A representative event is chosen from each different event type as an event template.

In real-time phase, real time video event features are again extracted following the above mentioned methodology. Extracted features are compared against all event templates generated in the training phase. Identity of the event to which the observation matches most is assigned to the observed event. If disparity between observed event and all templates exceeds a predetermined threshold T_D, that event is classified as an anomaly.

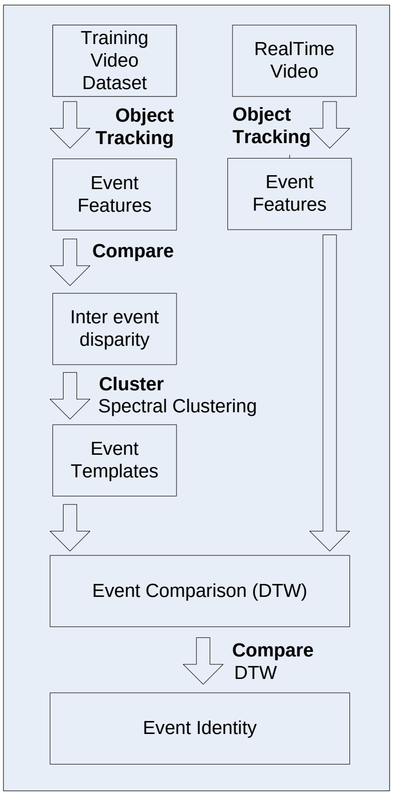

Figure1: Overview of the event classification and abnormal event detection methodology

3. EVENT DISPARITY EVALUATION

3.1. Temporal Feature Trajectory Comparison

In event classification, events are classified based on their similarity, evaluated in terms of features which describe them. More specifically the temporal feature trajectories of events are compared. The selections of the candidate temporal features are based on the desired classification outcome. For instance in a particular surveillance application the focus might be in the motion patterns of objects, where a feature such as object location trajectory could be used in event similarity evaluation.

When comparing two temporal feature trajectories, use of conventional signal comparison processes like correlation, direct mean square error or signal statistics are inappropriate since the length of compared trajectories would be different. Any effort for normalization through temporal axis of the feature trajectories to equalize the lengths might corrupt the trajectory information. This brings up the necessity of using more specific trajectory comparison methodologies for event detection applications. A utilized method for event similarity evaluation should reflect a concise scalar similarity between events when adherent for spectral clustering applications.

In catering the above problem of different length feature trajectory comparison Dynamic Time Warping (DTW) is employed.

3.2. Event Disparity through Dynamic Time Warping (DTW)

DTW is a specifically crafted algorithm for the purpose of different length feature trajectory comparison. The prominence given by the algorithm towards the shape and placement similarity of the compared two trajectories makes it useful in evaluating the motion pattern disparity of objects in a video scene. Though the classical DTW algorithm is based on single dimensional signal comparison, the methodology extends it resolving strengths with the expansion across dimensions.

The algorithm as in [5] defines a cost matrix in which elements define the cost of temporal alignment between the compared two trajectories. In adapting the temporal alignment matrix for the two dimension trajectory comparison of object motion patterns the cost matrix is defined as in,

$$C(N - i, j) = \left|\overleftarrow{T_1(\iota)} - \overleftarrow{T_2(\jmath)}\right|^2 \qquad (01)$$

Where $T_n(i)$ is the location vector nth object in the ith temporal instant and N is the length of trajectory placed parallel to a matrix column. Illustrated in Figure 2 is the constructed cost matrix for two such temporal object trajectories placed by the axes of the cost matrix image.

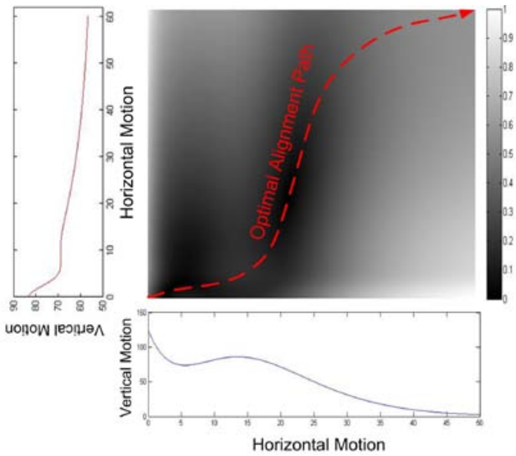

Figure 2: Alignment cost matrix in DTW and the placement of the optimal alignment path

Based on constructed matrix C, DTW algorithm seeks the optimal cost path from the beginning of the trajectories (bottom-left corner) to the termination end (top-right corner). As an additional constraint the optimum path is not allowed to propagate backwards temporally. Hereafter, the normalized event disparity $||E_i - E_k||$, between the events i and k are evaluated as in,

$$||E_i - E_k|| = {\{\sum_{(n.m)} C(n,m)\}}\Big/{\sqrt{N^2 + M^2}} \qquad (02)$$

Where,$(n, m) \in Optimal\ Cost\ Path$ and N, M are the lengths of the two trajectories being compared. Moreover, the inherent capacity of the DTW algorithm to remove certain non-linear variations in the temporal domain between the trajectories makes it useful in computing the mutual shape disparity between features.

4. EVENT CLASSIFICATION ON LEARNING SET

Forbye to classify any appearing video event, a learning set is defined and evaluated for its event context. In order to preserve the unsupervised event learning process intact spectral clustering is exploited for this purpose.

4.1. Spectral Clustering (SC)

Spectral clustering refers to a class of unsupervised techniques relying on the Eigen structure of a similarity matrix in segregation of event points into disjoint clusters with points in the same cluster having high similarity and points in different clusters having low similarity. Similarly for a given event network, it enables to find out K number of well separated clusters with information of component events in each cluster. The main steps of the algorithm are as following forbye to the version of normalized spectral clustering proposed in [6]:

1. Initially formed affinity matrix $A \in R^{n \times n}$ defined by,

$$A(i,j) = e^{\frac{-D_{i,j}^2}{2\sigma^2}} \qquad (03)$$

Where A is positive, symmetric and depends on the event disparity $D_{i,j}$ between the event data points.

2. Defined D to be diagonal matrix whose (*i*,*i*) element is the sum of A's ith row, and construct the normalized matrix,

$$L = D^{-\frac{1}{2}} A\, D^{-\frac{1}{2}} \qquad (04)$$

3. Eigen vectors $x_1, x_2, \dots, x_k$ corresponding to largest *K* Eigen values (choosing to be orthogonal to each other in the case of repeated Eigen values) of L are stacked as columns to form the matrix *X* as in, $X = [x_1, x_2, \dots, x_k] \in R^{n \times k}$.

4. Formed the matrix Y by renormalizing each row of matrix X to have a unit length according to,

$$Y_{ij} = \frac{X_{ij}}{\sqrt{\sum_j X_{ij}^2}} \qquad (05)$$

5. By treating each row of Y as a point in a R^n cluster space each row is clustered into K clusters via K-means algorithm by minimizing the distortion.

Here, the scaling parameter σ^2 controls how rapidly the affinity $A(i,j)$ fall off with the Euclidian distance $D_{i,j}$ between the event data points and in next part will describe a method for choosing it automatically by identifying K number of clusters. Further, to proceed K-means clustering the number of clusters K should be known in advance.

4.2. Initial Parameter Selection for SC

The dependency by the parameters of K and σ for performance in spectral clustering is enormous. By selecting proper parameters can ensure well separated event clusters in cluster space.

Even though there are verity of methods available to determine parameter K [7, 3], selection of parameter σ is completely disregarded in most work. In general, since there is strong dependency between K and σ, selecting the number of clusters K by disregarding the parameter σ happen to be completely in vain.

Further, conformity score based mechanisms are proposed in [7, 3] to determine K. However, with the variation of parameter σ, the generated cluster event classes might not represent actual event classes by the algorithm. Therefore separating actual event classes through cluster diagram is impossible, even though it can obtained an accurate estimate for K. To overcome this problem, a more generic method is proposed to determine K by considering the effect of σ, based on intrinsic properties of Laplacian matrix.

Referring to the matrix perturbation theory [8], perfect clustering in a dataset with K clusters is obtained when Eigen gap between Eigen values X_K^{th} and X_{K+1}^{st} are higher than others. While scanning through a search space of σ, it can be seen that the differences between consecutive Eigen values are varying in well-defined pattern. Another noticeable fact is that the first Eigen difference to reach a given threshold value (in our observation 0.2) is always the difference between X_K^{th} and X_{K+1}^{st} Eigen values. This observation is consistent irrespective of the selected dataset.

Moreover, considering this aspect, minimum value of σ that causes Eigen gap between any two consecutive Eigen values X_N and X_{N+1} to go beyond a threshold *T=0.2* is selected as optimum σ and number of clusters K is taken to be equal to N as in Figure 3.

5. EVENT CLASSIFICATION AND ANOMALY DETECTION

In identifying the event class of any emerging event, their disparity is evaluated with template representatives of the learning set event classes at set intervals. A template representative from each event class is selected such that the sum of disparities from the template representative to other co-class events is minimum as in,

$$E_{T,J} = Argmin_{E_i}\left\{\sum_{k=1}^{N_J}\|E_i - E_k\|\right\} \qquad (06)$$

Where, T is the selected template event for a class J with N_J cases and E_i represents i th event of the class J.

Each of the events is attributed to the event class having the minimum disparity figure from the representative event template. Additionally, a pre-determined disparity threshold, $T_D$marks the maximum bound of disparity for an activity to be considered as normal. The overall event classification process could be summarized as in,

$$I = \begin{cases} Abnormal & ; \ \|E_i - E_{T,J}\| > T_D \ for \ \forall \ J \\ Argmin_J \|E_i - E_{T,J}\| & ; \ otherwise \end{cases} \qquad (07)$$

Where E_i is the observed event and $E_{T,J}$is the template representative for the class J. In order to determine threshold T_D, the event disparities between each of the events with the class

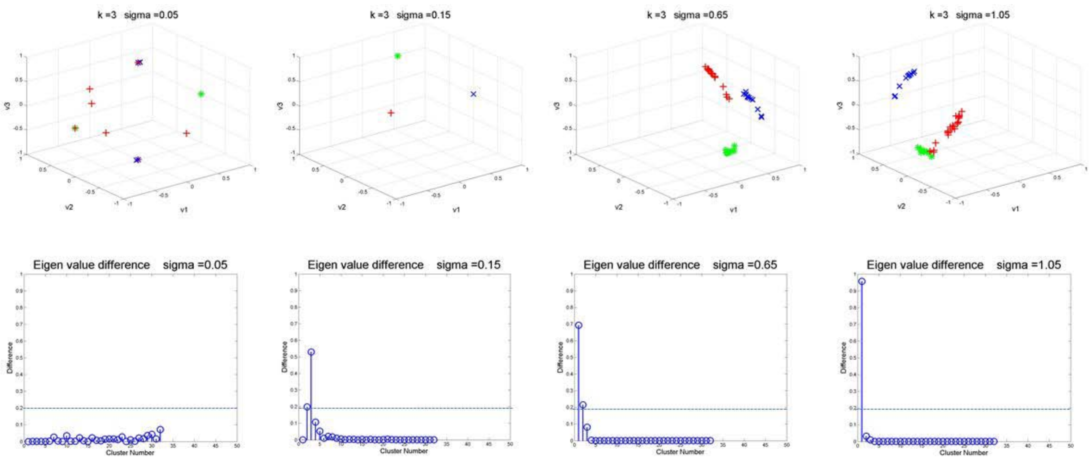

Figure 3: Parameter optimization process through eigen analysis of the laplacial matrix

representatives are evaluated constructing the disparity histogram shown in Figure. T_D is defined as a value separating the two identified distance categories "Disparity from co-class representatives" and "Disparity from other representatives". In this application T_D is chosen to be 0.05.

6. RESULTS

The overall event comparison and classification process was implemented and tested for real life video events focusing on human motion patterns. The test video set contained 90 object motions belonging to 5 major event classes and 15 anomalous events. Illustrated in Figure 5 below are some of the identified cases through the methodology in which the event class is signified from the respective color and the identity number.

In preprocess of this methodology, a random selection of 50 events was imposed as the learning set and rest of the cases were tested for classification results individually. The methodology was able to classify 92% of the above cases accurately for fully available motion trajectories.

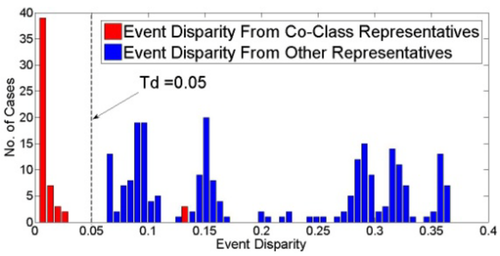

Figure 4: Selection of disparity threshold (T_D) for abnormal activity detection through disparity histogram

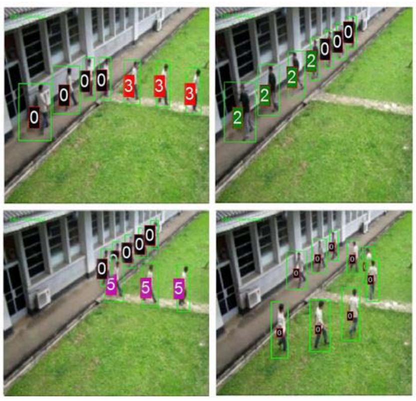

Figure 5: Few of the identified cases through the overall process

In each of learning set clustering instances the automated cluster parameter determination method was utilized and able to find out number of clusters with respect to optimum sigma value. Abnormal activities that differ from standard event templates can be identified using the defined threshold (T_D) with a value of 0.05 and this process was able to identified all the 15 anomalous motion in the test data successfully. The threshold (T_D) value was determined considering the observations for matching and non-matching class distances with the test point.

7. CONCLUSION

A methodology for video event detection through feature trajectory comparison is presented in this paper. DTW is employed in comparing the available feature trajectories pairwise through a mutual disparity figure. Utilized spectral clustering algorithm and introduced initial parameter selection enables in identifying event class from the available learning set preserving unsupervised learning. Comparison with selected class representatives from the learning events is carried out in classifying any emerging event or deciding as an abnormal event. Results have been demonstrated for a series of real life videos with regard to human motion patterns.

8. REFERENCES

[1] H. M. S. P. B. Herath, W. S. K. Fernando, P. H. Perera, M. P. B. Ekanayake, G. M. R. I. Godaliyadda, and J. V. Wijayakulasooriya, "*Human motion tracking under dynamic background conditions*", in: Proceedings of 9th IEEE International conference on industrial and information systems (ICIIS), Gwalior , India, pp. 1-6, 2014.

[2] F. Jiang, Y. Wu, and A. K. Katasggelos, "*Detecting contextual anomalies of crowd motion in surveillance video*", in: proceedings of 16th IEEE International conference on image processing (ICIP), Cairo, pp. 1117-1120, 2009.

[3] F. Porikli, and T. Haga, "*Event detection by eigen vector decomposition using object and frame features*", in: Proceedings of Computer vision and pattern recognition workshop, CVPRW, pp 114, 2004.

[4] A. Wiliem, V. Madasu, W. Boles, and P. Yarlagadda, "*Detecting uncommon trajectories*", in Proceedings of International conference on digital image computing: Techniques and applications, Canberra ACT, pp. 398-404, 2008.

[5] M. Muller, "*Dynamic time warping*",

Information retrieval for music and motion Part I, Springer Berlin Heidelberg. pp. 69-84,2007.

[6] A. Ng, M. Jordan, and Y. Weiss, "*On spectral clustering: Analysis and an algorithm*", Advances in Neural information processing systems (eds. Dietterich, T., Becker, S.and Ghahramani, Z.), MIT Press, Cambridge, pp. 849-856, 2002.

[7] L. Hao, D. Bull, and A. Achim, "*Unsupervised video anomaly detection using feature clustering*", IET Signal Processing, Vol. 6, pp. 521-533, 2012.

[8] U. von Luxburg, "*A tutorial on spectral clustering*", Jouranl of Statistics and Computing, Vol 17, pp. 395-416, 2007.