International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1, October, 2019

Wasserstein Clustering based Video Anomaly Detection for Traffic Surveillance

S Arivazhagan, M Mary Rosaline, W Sylvia Lilly Jebarani

Abstract*: Anomaly Detection is very important in present scenario with huge availability of data and enormous difficulty in extraction of meaningful information out of it. In this paper we present an approach for video anomaly detection based on trajectory features and spatio – temporal features. Clustering of spatio – temporal features and trajectory features are performed in Wasserstein metric space and cluster distance and span in Wasserstein metric space is exploited to perform anomaly detection. The Performance of the Anomaly detection with Wasserstein distance based K – means and Wasserstein distance based DBSCAN clustering of the 3D wavelet features and trajectory features was studied. The method is robust and suffers from fewer false alarms.

***Keywords* : Wasserstein Distance, Anomaly Detection.**

I. INTRODUCTION

The anomaly detection is essential in Video Surveillance, Remote Sensing, Medical Diagnosis and Treatment, Eco system monitoring, Driver Assistance systems, Cyber-Intrusion Detection, Medical Anomaly Detection, Machinery fault Detection, Forest and crop monitoring, Drought and flood monitoring , Land slide and Earth quake detection. Video Anomaly detection is often approached as trajectory anomaly detection. The trajectory anomaly detection is performed by extraction of the trajectories of the moving objects in the video and clustering of the trajectories [1]. The features normally used for trajectory clustering include trajectory curvature, motion histogram etc. The largest cluster with most members is considered to be of normal and any trajectory outside of the cluster is detected to be an anomaly. K – Means clustering is a popular clustering method when the number of clusters is known. The other clustering strategies include sequential clustering, hierarchical clustering [3], spectral clustering [2], mean shift clustering, Kernel Density Estimation (KDE) clustering etc. The disadvantage with trajectory based detection strategy is that the trajectory detection is not always possible. The K – means algorithm is preferred for trajectory clustering when the number of clusters can be estimated, whereas algorithms like spectral clustering, hierarchical clustering, mixture model, mean shift, SVM, and kernel density estimation are preferred otherwise.

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
S Arivazhagan, Department of ECE, Mepco Schlenk Engineering College, Sivakasi, India.Email:sarivu@mepcoeng.ac.in
M Mary Rosaline*, Department of ECE, Mepco Schlenk Engineering College, Sivakasi, India. Email :maryrosaline@mepcoeng.ac.in
W Sylvia LillyJebarani, Department of ECE, Mepco Schlenk Engineering College, Sivakasi, India.Email:wsylvia@mepcoeng.ac.in

Apart from trajectory based anomaly detection, spatio –temporal feature based anomaly detection is also quite popular either at pixel level or at patches level. Zhong et al. [4] proposed representation of video motion with the help of spatiotemporal gradients of pixels and spectral clustering of the gradients for anomaly detection. Boiman and Irani [5] proposed clustering of spatio-temporal patches and event modeling with cluster representation, with largest cluster representing normality and other clusters deviated from it representing abnormality.

Event based video anomaly detection is another class of anomaly detection, performed at higher level based on event clustering. Hamid et al. [6] proposed representation of activities as bags of event n-grams, and analyzing the global structural information of activities using local event statistics. Wang et al. [7] represented video events as distributions over low-level visual features on a pixel basis and used hierarchical Bayesian models for event clustering. By considering temporal context, an anomalous video event may include behaviors at multiple times, i.e., having an arbitrary length of time. Many of the existing works fail to provide any modeling of such contextual information

The Wasserstein distance is an optimal metric for distance measures [8, 10], the practice of which is strongly supported by optimal transport problem. In this paper we propose Wasserstein distance based trajectory feature clustering and spatio – temporal features clustering for anomaly detection.

The Section 2 discusses about the Wasserstein clustering based anomaly detection. Section 3 and the Section 4 explain about Wasserstein distance based K – means clustering and Wasserstein distance based DBSCAN clustering respectively. The Section 5 gives the list of anomaly measures used, while the Section 6 and the Section 7 elaborate on the experimental results and conclusion respectively.

II. WASSERSTEIN CLUSTERING BASED ANOMALY DETECTION

A. System Methodology

The Proposed system architecture for Anomaly detection is given in the Fig. 1. The features used for the anomaly detection system are spatio temporal 3D Wavelet features for video anomaly detection and trajectory FFT features for trajectory anomaly detection. The features are extracted from the reference data set using feature extraction system. Then the features are clustered using the proposed Wasserstein distance based K – Means / BSCAN clustering.

These reference clusters and their Wasserstein Barycenters are stored in a database.

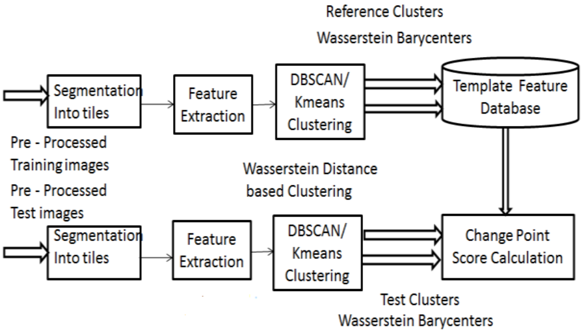

Fig. 1. Anomaly Detection System Methodology

The same set of features is extracted from the test data set using feature extraction system. Then the features are clustered using Wasserstein distance based K – Means / DBSCAN clustering. These test and reference clusters and their Wasserstein Barycenters are given to the anomaly detection system which calculates the anomaly score point. K – Means clustering is preferred when the number of classes in the training data is known in prior and DBSCAN clustering is preferred when the number of classes in the training data is not known in prior. The features needed for Wasserstein distance based anomaly detection include set valued feature vectors based on 3D wavelet and FFT features. The anomaly detection can exploit only the spatial features if the video environment is static in nature. The extraction of spatio - temporal features is necessary when the environment is dynamic. In spatial feature extraction, 2D Wavelet and FFT features are extracted from spatial grid points and in spatio – temporal feature extraction, 3D Wavelet and FFT based features are extracted from spatio - temporal grid points as given in Fig. 2.

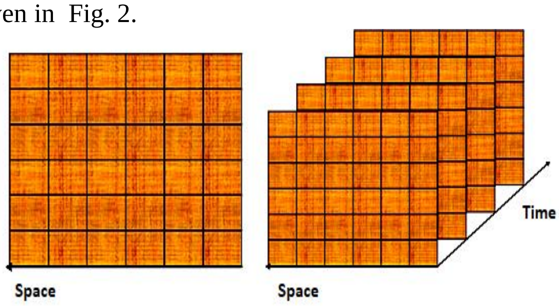

Fig. 2. Spatial and Spatio temporal grid points

B. Wasserstein Distance

For comparing set- valued data like Histogram, Fourier coefficients and Wavelet co-efficient, Euclidean distance and Manhattan distance do not suffice. In such cases Wasserstein Distance is the best criteria. The Wasserstein distance between the two probability distributions given in the Fig. 3. can be explained as the cost of turning Distribution A to Distribution B. To Change Distribution (sand heap) A to Distribution (sand heap) B some portion from the peak of the Distribution (sand heap) A should be moved to the left and right corner of it. Wasserstein distance between two distributions is the cost of turning one distribution into another. Wasserstein distance (or Earth Movers distance) [12, 13] can be defined as the amount of Points (Sand Particles) moved multiplied by the distance they are moved. The points (Sand Particles) should be moved in the optimum possible way.

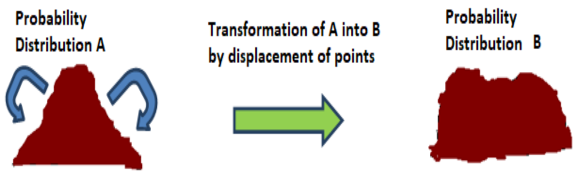

Fig. 3. Wasserstein Distance Illustration

The amount of points and the distances they have to be moved should be selected in such a way that it would result in closest match possible between them and it would take least possible movement of the points to attain the same. The N – Wasserstein distance can be written as given in (1),

where P and Q are probability distributions, F_{ij} is the

$$W_N(p,q) = \left(\inf_{F} \sum_{i=1}^{m}\sum_{j=1}^{n} F_{ij}\, d(x_i,y_j)^N\right)^{\frac{1}{N}}$$

$$\sum_j f_{ij} \le P_i, \sum_i f_{ij} \le Q_j,$$

$$\sum_{i,j} f_{ij} = \min\left(\sum_i P_i, \sum_j Q_j\right), f_{ij} \ge 0 \qquad (1)$$

probability that a point in i^{th} bin has fallen on j^{th} bin or the number of points that have to be moved from bin *'i'* to bin *'j'* and $d(x_i,y_j)$ the distance between i^{th} bin from which the point has to be moved and j^{th} bin to which the point has to be moved.

This problem can be solved by solving Linear Programming transportation problem.

III. WASSERSTEIN DISTANCE BASED K – MEANS CLUSTERING

Proposed Wasserstein Distance based *K* – Means clustering is similar to *K* – means clustering and differs only in the fact that data is clustered into *K* clusters such that sum of mean of the square of the Wasserstein distance of points to the centers of all the clusters is minimum. Whereas, in *K* – means clustering, data is clustered into *K* clusters such that sum of mean of the square of the Euclidean distance of points to the centers of all the clusters is minimum.

In other words, in *K* – means clustering the points are clustered into *K* clusters in such a way that each point is associated to the closest centroid in Euclidean space where as, in Wasserstein Distance based *K* – means clustering the points are clustered into *K* clusters in such a way that each point is associated to the closest centroid in Wasserstein metric space. Wasserstein Barycenter of the cluster points is chosen as the centroid of the cluster in Wasserstein metric space. The flow chart for the Proposed Wasserstein Distance based *K* – Means clustering is shown in the Fig. 4.

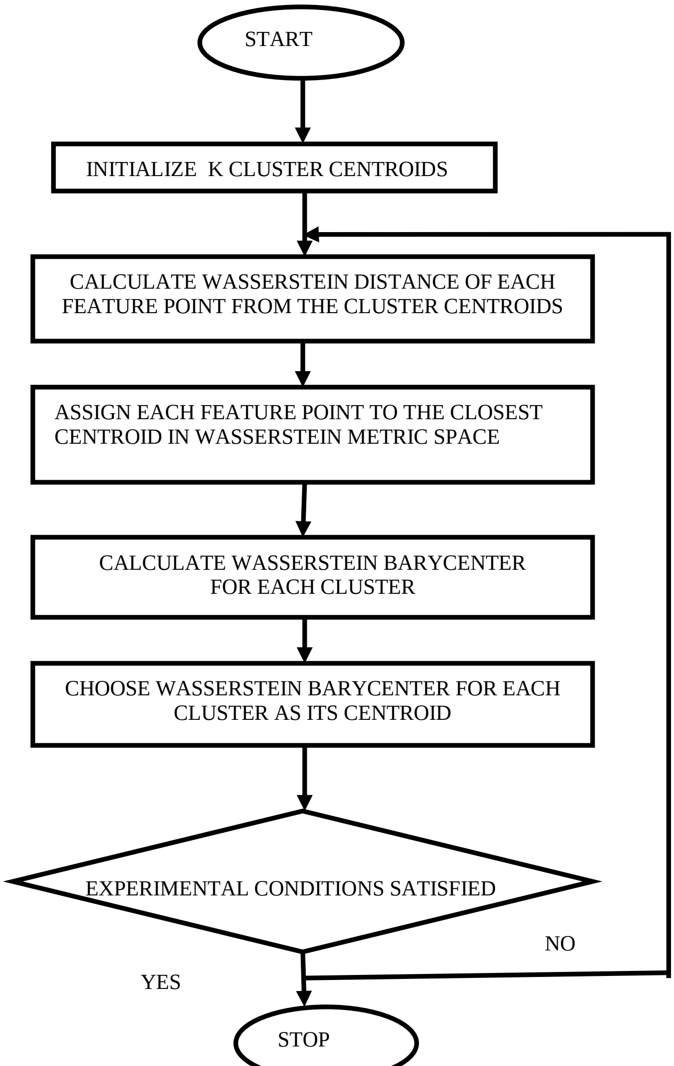

Fig. 4. Wasserstein Distance based K – means Clustering

IV. WASSERSTEIN DISTANCE BASED DBSCAN CLUSTERING

DBSCAN clustering is density based clustering where the points are clustered in such a way that a minimum density of points is maintained throughout the region within the cluster with density being the number of points per unit radius. A point is included to a cluster only when the number of points within the preselected Euclidean or Manhattan distance from the point is greater than a minimum threshold.

The proposed Wasserstein Distance based DBSCAN clustering clusters the data such that each point would have minimum number of w - Wasserstein neighbors. w - Wasserstein neighbors of a given point are the points that are at Wasserstein distance of less than 'w' from the given point. The border points will have fewer neighbors compared to points inside the cluster. The flow chart for the Proposed Wasserstein Distance based K – Means clustering is shown in the Fig. 5.

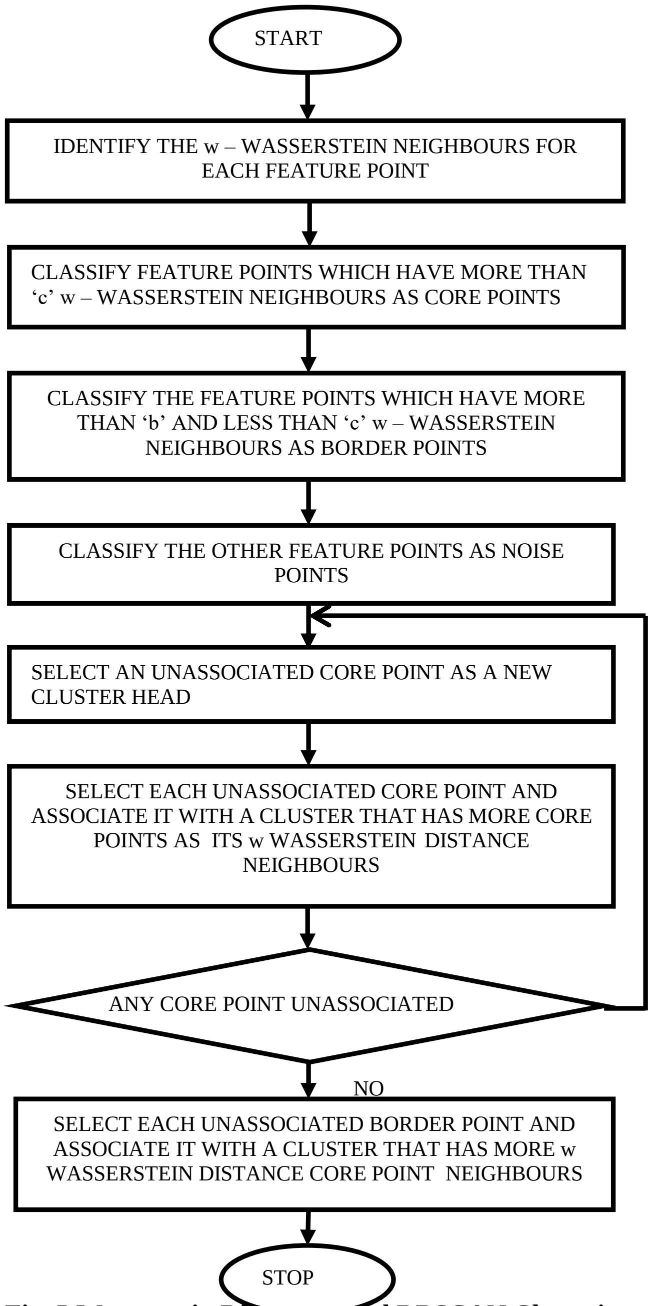

Fig. 5.Wasserstein Distance based DBSCAN Clustering

V. ANOMALY SCORE POINT CALCULATION

The first step is embedding the features in an appropriate metric space with a distance measure between the features. The next step is measuring anomaly based on the distance between the reference feature set and the test feature set. The anomaly is detected based on the location and the span of the test feature clusters with respect to the location and the span of the reference feature clusters. We propose the following Anomaly Score Points on Wasserstein metric space.

A. Log Density Ratio

The Log Density Ratio based Anomaly Score Point is a distance measure based on entropy estimators. The log density ratio score is the measure of distance between the reference cluster R and test cluster T and can be given by (3)

$$Score(T) = I(B;R) - I(B;T) \quad (3)$$

Where *I(B;R)* is the exclusive information in Reference cluster with respect to Reference Wasserstein Barycenter , *I(B;T)* the exclusive information in Test cluster with respect to Reference Wasserstein Barycenter and *B* the Wasserstein Barycenter of reference cluster. Information in a cluster *R* with respect to barycenter *B*, *I(B;R)* can be given by (4).

$$I(B;R) = c + d \sum_{B_i \in R} w_i \log D(B_i, B) \quad (4)$$

Where $D(B_i,B)$ is the Wasserstein distance between i^{th} feature in the cluster and *B* , B_i the i^{th} feature in the cluster and *B* the considered reference Wasserstein barycenter.

B. Symmetrized Kullback – Leibler Divergence

The Symmetrized Kullback – Leibler Divergence based Anomaly Score Point can be given by (5)

$$Score(T) = \frac{2H(R,T) - H(R) - H(T)}{2} \quad (5)$$

where *H(R,T)* is the cross entropy between the reference cluster and test cluster and *H(R)* and *H(T)* the auto entropy of the reference and test clusters. The *H(R,T)* the cross entropy between *R* and *T* can be given by (6)

$$H(R,T) = c + d \sum_{\substack{R_i \in R \\ T_j \in T}} \frac{w_i w_j}{1-w_i} \log D(R_i, T_j) \quad (6)$$

The *H(R)* the auto entropy of *R* can be given by (7).

$$H(R) = c + d \sum_{\substack{R_i, R_j \in R \\ i \neq j}} \frac{w_i w_j}{1-w_i} \log D(R_i, R_j) \quad (7)$$

where *c* and *d* are constants.

C. Mean Square Wasserstein Distance

The Mean Square Wasserstein Distance based Anomaly Score Point can be defined as the difference between the mean square Wasserstein distance of the reference cluster from the reference Wasserstein barycenter and mean square Wasserstein distance of the test cluster from the reference Wasserstein barycenter and is given by (8)

$$Score(T) = D(B;R) - D(B;T) \quad (8)$$

Where.

$$D(A;B) = \sum_{i=1}^{n} w_i W_2^2(A, B_i) \quad (9)$$

Where w_i are the weights for weighted mean operation, $W_2^2(A,B_i)$ is the square of the 2 - Wasserstein distance between i^{th} feature in the cluster and its barycenter , B_i the i^{th} feature in the cluster and *A* its Wasserstein barycenter.

D. Modified Weighted mean Variance Distance

The Modified Weighted mean Variance Distance based Anomaly Score Point can be given by the sum of the difference between the mean square Wasserstein distance of the reference cluster from the reference Wasserstein barycenter and mean square Wasserstein distance of the test cluster from the test Wasserstein barycenter and the difference between the reference barycenter and the test barycenter as given in (9)

$$Score(T) = D(B;R) - D(B;T) + W_R - W_T \quad (10)$$

Where.

$$D(A;B) = \sum_{i=1}^{n} w_i W_2^2(A, B_i) \quad (11)$$

Where w_i are the weights for weighted mean operation, $W_2^2(A,B_i)$ is the square of the 2 - Wasserstein distance between i^{th} feature in the cluster and its barycenter , B_i the i^{th} feature in the cluster and A its Wasserstein barycenter. W_R is the reference cluster barycenter and W_T the test cluster barycenter.

E. Nearest Neighborhood distance

The Anomaly Score Point of the test feature is the Wasserstein distance to the closest feature in the reference cluster. It is given by (10).

$$Score(T) = \min_{i} W(T, B_i) \quad (12)$$

VI. EXPERIMENTAL RESULTS AND DISCUSSION

Our proposed method is implemented using MATLAB 2014b. The method is tested in an Intel core i5 750 PC machine with a image sequence extracted from PETS 2009 and a UCSD dataset. For the construction of feature vectors, we partition the video frames into blocks of 15 frames each. 3D Wavelet features are extracted from the blocks for both training and test datasets

The 3D wavelet features are clustered in the Wasserstein metric space with the help of proposed Wasserstein distance based K means / DBSCAN clustering algorithms for both training and test data sets. The anomaly score point is calculated based on comparison of cluster span and position of clusters of the training and test dataset.

In UCSD dataset the abnormal movements of a speeding motorcyclist, skatist, wheel chair, person walking on grass lane in the road are detected. The Fig. 6.a. shows the anomaly score for the abnormal movement of a speeding motorcyclist in the road. The Fig. 6.b. shows sample frames of the detected anomaly in the UCSD dataset. The Fig. 6.c. shows the anomaly score for the abnormal movement of a person skating speedily in the road whereas the Fig. 6.d. shows sample frames of the detected skating anomaly in the UCSD dataset.

The proposed method is tested for trajectory dataset. For the construction of feature vectors, we took DFT on the trajectory data. The DFT features are clustered in the Wasserstein metric space with the help of proposed Wasserstein distance based K means / DBSCAN clustering algorithms for both training and test data sets. The anomaly score point is calculated based on comparison of cluster span and position of clusters of the training and test dataset. In trajectory dataset the abnormal movement of the vehicles in the road is detected. Fig. 6e shows the clustering of normal trajectories in the Wasserstein metric space and Fig. 6f show the anomaly detection result for the trajectories, with red colour trajectories being the anomalous trajectories.

Retrieval Number: A2222109119/2019©BEIESP
DOI: 10.35940/ijeat.A2222.109119
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table- I: Anomaly Detection Results for UCSD for Pedestrian Dataset

Type of Anomaly	Number of Frames	Detection Rate	False Alarm Rate
Skating	982	100	0
Bi - cyclist	2435	100	0
Wheel Chair	335	100	0
Car / Truck	430	100	0
Cart	130	100	0
Waking on grass lane	170	100	0

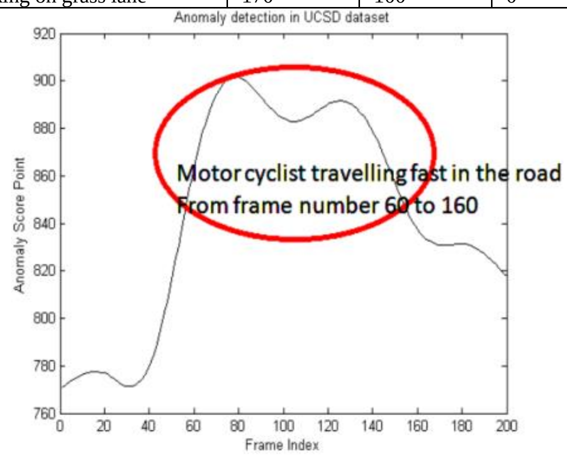

Fig. 6.a.

Fig. 6.b.

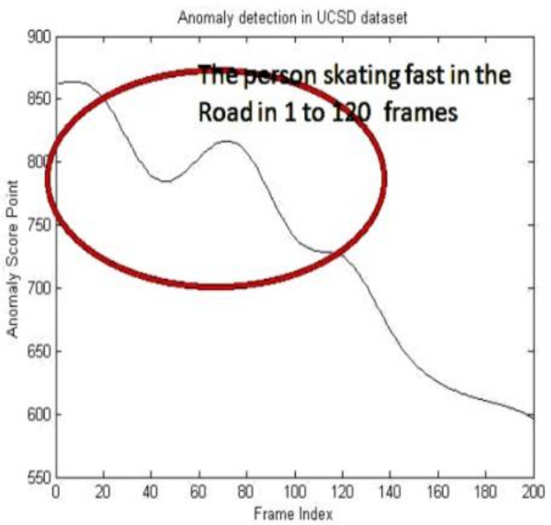

Fig. 6.c.

Fig. 6.d.

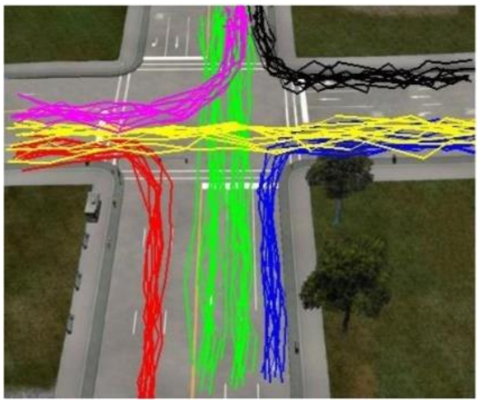

Fig. 6.e.

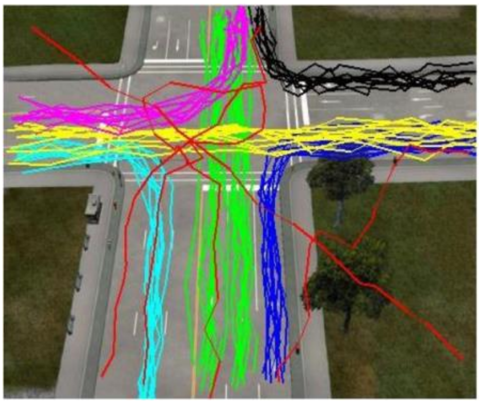

Fig. 6.f.

Fig. 6. a. Anomaly Score point value for speeding cyclist b. Speeding cyclist travelling in the road c. Anomaly Score point value for skatist d. Skatist travelling in the road e. Wasserstein distance based trajectory clustering f. Wasserstein Clustering based trajectory anomaly detection.

The performance of the Anomaly detection Algorithm for UCSD pedestrian dataset is summarized in the Table I. The performance of Wasserstein clustering based anomaly detection is compared with Euclidean Clustering based anomaly detection and is found to be superior in terms of accuracy percentage. With Euclidean Clustering based anomaly detection the accuracy was about 82%.

Retrieval Number: A2222109119/2019©BEIESP
DOI: 10.35940/ijeat.A2222.109119
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

VII. CONCLUSION

In this work clustering in Wasserstein metric space was explored. The application of the same was extended to Video Anomaly detection. The performance of Wasserstein distance based K – means clustering and DB – SCAN clustering was evaluated for anomaly detection in UCSD and PETS pedestrian dataset and trajectory dataset. Trajectory anomaly detection was also evaluated for trajectory dataset. Wasserstein clustering is found to be effective compared to Euclidean distance and Manhattan distance based clustering for set-valued features.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Management, Principal and Head of the Department of Electronics and Communication Engineering of our college for their constant support and encouragement.

REFERENCES

1. C.R. Jung, L. Hennemann, S.R. Musse, "Event detection using trajectory clustering and 4-d histograms", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, issue. 11, 2008, pp. 1565–1575.
2. F. Porikli, T. Haga, "Event detection by eigenvector decomposition using object and frame features", in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, vol. 7, 2004, pp. 114–124.
3. F. Jiang, Y. Wu, A.K. Katsaggelos, "A dynamic hierarchical clustering method for trajectory-based unusual video event detection", *IEEE Transaction on Image Processing*, vol. 18, issue. 4,2009. pp. 907–913.
4. H. Zhong, J. Shi, M. Visontai, "Detecting unusual activity in video" *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* , vol. 2, 2004, pp. 819–826.
5. O. Boiman, M. Irani, "Detecting irregularities in images and in video" *Proceedings of IEEE Int'l Conference on Computer Vision*, vol. 1, 2005, pp. 462–469.
6. R. Hamid, A. Johnson, S. Batta, A. Bobick, C. Isbell, G. Coleman, "Detection and explanation of anomalous activities: representing activities as bags of event ngrams", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, (vol. 1), 2005, pp. 1031–1038.
7. X. Wang, X. Ma, W.E.L. Grimson, "Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, issue. 3, 2009, pp. 539–555.
8. Y. Rubner, C. Tomasi, and L. J. Guibas, "The Earth mover's distance as a metric for image retrieval," *Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.
9. J. Wang, M. Baum, P. Willett, and U. D. Hanebeck, "On Wasserstein barycenters and MMOSPA estimation," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1511–1515, Oct. 2015.
10. O. Pele and M. Werman, "Fast and robust Earth mover's distances," *Proceedings of International Conference on Computer Vision*, 2009, pp. 460–467.
11. E. Anderes, S. Borgwardt, and J. Miller, "Discrete Wasserstein barycenters: Optimal transport for discrete data," Mathematical m Methods on Operation Research, vol. 84, no. 2, pp. 389–409, 2016..
12. Yasen Zhang, Xian Sun, Hongqi Wang, and Kun Fu, "High-Resolution Remote-Sensing Image Classification via an Approximate Earth Mover's Distance-Based Bag-of-Features Model", *IEEE Geoscience and Remote Sensing Letters*, vol. 10, issue. 5, 2013, pp. 1055 - 1059.
13. Rubner. Y.,Tomasi. C., and Guibas. L., "The Earth Mover's Distance as a Metric for Image Retrieval", *International Journal of Computer Vision*, vol. 40, issue. 2, 2000, pp. 99–121.
14. Jain A.K., Zhong. Y., and Lakshmanan S., "Object Matching Using Deformable Templates", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, issue. 3, 1996, pp. 267-278.
15. Greenspan H.,Dvir G., and Rubner Y., "Region Correspondence for Image Matching via EMD Flow", *Proceedings IEEE Workshop Content- Based Access of Image and Video Libraries*, 2000
16. Peihua Li, Qilong Wang, Lei Zhang, "A Novel Earth Mover's Distance Methodology for Image Matching with Gaussian Mixture Models", *IEEE International Conference on Computer Vision* ,2013.
17. F. Memoli, "On the use of Gromov-Hausdorff distances for shape comparison", *Proceedings of Point Based Graphics*, 2007.
18. F. Memoli, "Gromov-Hausdorff distances in Euclidean spaces", *Proceedings on IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008, pp. 1–8.
19. Kensuke Koshijima, Hideitsu Hino, and Noboru Murata , "Change-Point Detection in a Sequence of Bags-of-Data", *IEEE transactions on Knowledge and Data Engineering*, vol. 27, issue. 10, 2015, 2632 - 2644.

AUTHORS PROFILE

S. Arivazhagan received his B.E degree in Electronics and Communication Engineering, from Alagappa Chettiar College of Engineering and Technology, Karaikudi in 1986 .He completed his M.E. degree in Applied Electronics in College of Engineering, Guindy, Anna University, Chennai in 1992. He completed his Doctorate in Texture Classification using Wavelet transform in the year 2005. Currently, He is working as the Principal in Mepco Schlenk Engineering College, Sivakasi. He has published more than 170 Technical papers in the International /National Journals and Conferences. His current research interests include: Image processing, pattern recognition and computer communication.

M. Mary Rosaline is currently working as Assistant Professor, in Department of Electronics and Communication Engineering, Mepco Schlenk Engineering College, Sivakasi. She received her B.E degree in the faculty of Electronics and Communication Engineering from Madurai Kamaraj University in 2003 and M.E. degree in Communication Systems from Anna University in 2005. Her current research interest is SAR Image Analytics.

W. Sylvia Lilly Jebarani received the B.E degree in Electronics and Communication Engineering from Mepco Schlenk Engineering College, M.K. University in 1998 and M.E degree in Communication Systems from Mepco Schlenk Engineering College, M.K. University in 2003. She has been awarded with Ph.D. degree by Anna University, Chennai in the year 2017. She is currently an Associate Professor, ECE Department of Mepco Schlenk Engineering College, Sivakasi. She has eighteen years of teaching and research experience. She has published around 69 Technical papers in International / National Journals and Conferences. Her current research interests include Digital Image Processing, Steganography, Steganalysis, and Computer and Communication Networking. She has completed five Research and Development Projects, funded by ADE, Bengaluru, NPOL, Kochi and DRDO, New Delhi.