Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 167 (2020) 890–899

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2019)

A Novel Approach for Pose Invariant Face Recognition in Surveillance Videos

Manju D[a]* and Radha V[b]

[a]Research Scholar, Department of Computer Science, Avinashilingam Institutre for Home Science and Higher Education for Women,(Assistant Professor, Depatment of Computing, Coimbatore Institute of Technology), Coimbatore-641043, India
[b]Department of Computer Science, Avinashilingam Institutre for Home Science and Higher Education for Women, Coimbatore-641043, India

Abstract

The face detection and face recognition methods are introduced to confirm the abnormal human activity in the video surveillance system. Face detection is carried out by Viola-Jones face detector. It is composed of three concepts namely integral image, AdaBoost and the cascade structure. After face detection, Histogram of Oriented Gradient(HOG) and Weighted Local Binary Patterns(WLBP) features are extracted and those are used in Orthogonal Locality Preserving Projection(OLPP) for face recognition. The detected faces may contain pose variation which dramatically degrades the OLPP based face recognition. So, pose-invariant OLLP-based face recognition is proposed where Histogram of Face Orientation(HFO) and Histogram of Face Direction(HFD), HOG and WLBP features are used in OLPP for efficient face recognition.

Keywords:Face detection; Face Recognition; Viola-Jones face detector; Histogram of Oriented Gradient; Weighted Local Binary Patterns;Orthogonal Locality Preserving Projection; Histogram of Face Orientation; Histogram of Face Direction.

* Corresponding author. Tel.: +91 98652 29679;
*E-mail address:*manju.kulasekar@gmail.com

10.1016/j.procs.2020.03.428

1. INTRODUCTION

Video surveillance has been given more attention in recent years in the computer vision community. Because of the rising demand for security and safety, more research has been conducted for intelligence-based surveillance systems. Watching people in Home-nursing, Shopping Malls, Education Institutions, Hospitals, waiting rooms and monitoring vehicles are few video surveillance applications. Nowadays, an intelligent system is integrated with video surveillance applications to analyze human behaviors automatically through the prediction of sequences of human activities. The prediction of human activity [1] is more significant to prevent transgression and treacherous activities from occurring.

The prediction of human activity is more crucial for the video surveillance system. Human activity is predicted by finding the probabilistic ratio of a specific activity to the overall activities in the video or finding associatively related events in the video. Human activity prediction methods in the video surveillance systems are used to find unrelated events or abnormal events from the normal events which already categorized and stored in the database. Many Human Activity prediction techniques were proposed to find activities from videos. In this paper, the human activity prediction is utilized to find the abnormal events from the surveillance videos.

A novel framework [2] was proposed for early prediction of human activities. A Probabilistic Suffix Tree (PST) was introduced in this framework that modeled the casual relationships between constituent actions. A Sequential Pattern Mining (SPM) was used to model the interactive objects information. A Predictive Accumulative Function (PAF) was presented to depict the predictability of each kind of activity. The prediction of human activity by this framework was improved by considering different information such as spatial, temporal, size of the blob and motion correlation among objects [3]. After the collection of information, a Spatio-Temporal Frequent Object Mining (STFOM) was applied to mine the frequent items (activities). The frequent activities were considered as normal activities and remaining activities were considered as abnormal activities.

Person identification is one of the most important aspects in video surveillance. To identify a person various features such as gait, clothes, posture but face has always been the most reliable features. In this paper, face detection and face recognition techniques are introduced to identify the person who is all involved in the abnormal activity. A Viola-Jones face detector is introduced for face detection where integral image, Ada Boost and the cascade structure concepts are used. The concept of Integral image evaluates the rectangular features in constant time. The Ada Boost learning algorithm selects the features and assigns a weight to all the selected features. Then a strong classifier is constructed as a linear combination of weak classifiers. In the cascade structure, combine the classifiers in a cascade manner. Then, the Histogram of Oriented Gradient (HOG) and Weighted Local Binary Patterns (WLBP) features are extracted from the detected face image and these features are used in Orthogonal Locality Preserving Projection (OLPP) for face recognition. The HOG feature represents the edge structure of the face and the WLBP feature capture the local texture information about the detected face. The detected faces might have the pose variations.

The above face detection and recognition methods are more effective only for frontal images. One of the key challenges in the current face recognition technique is posed variations. The recognition of faces from different poses is recognized by rotations of faces from various angles to the straight or in front angle. So, for recognition of faces at different poses, pose-invariant OLPP based face recognition is introduced. In pose-invariant OLPP based face recognition, Histogram of Face Orientation (HFO) and Histogram of Face Direction (HFD) features are used to define the orientation and direction of faces and it is used in OLPP for efficient face recognition. Hence, the proposed face detection and face recognition methods detect and recognize the faces more effectively in the video surveillance system. It prevents crimes and dangerous activities from occurring.

The organization of the paper is given as follows: Section 2 presents the literature survey related to the abnormal human activity prediction and face recognition. Section 3 explains the proposed methodology. Section 4 illustrates the experimental results of the proposed method. Finally, Section V concludes the research work.

2. Literature Survey

An entropy-based approach [4] was proposed for the detection of unusual events in the surveillance video. Entropy was a measure of randomness in a video frame. The entropy measured the spatiotemporal displacements of interested regions between video frames. The statistical calculation of entropy variation was calculated for each action in the video. If the statistical variation was deviated from the normal activities were found as abnormal activities. The normalized entropy value was used to estimate the activities for multi camera surveillance systems. However, the overlapping occurrences of multi-camera were the limitation of this work.

A context-aware approach [5] was proposed for abnormality prediction in ambient assisted living. In this approach, a Hidden Markov Model (HMM) was exploited to predict abnormalities and behavioral trends in various user contexts. HMM was utilized to obtain abnormality sequences from historical events in the database. An exponential smoothing approach was used to predict the future abnormality sequences. The results of variant models were combined by fuzzy logic to take final decision about the abnormality prediction. However, the fuzzy model requires more space to store the rules.

An abnormal activity recognition [6] approach was proposed based on the graph formulation of video activities. It also proposed the use of a graph kernel support vector machine. Initially, a graph was developed based on the interaction of the entities in a video. The nodes in the graph represented the entities and the vertices represented the spatiotemporal interest points. Then, a binary support vector machine with a graph kernel was applied in the graph to classify the activities as normal and abnormal activities. However, the support vector machine required extensive memory for classification.

A novel methodology [7] was proposed for the prediction of spatiotemporal activities in an urban environment. The latent characteristics representing the different temporal factors and spatial environments were determined and quantified through tensor factorization. This methodology assumed that the spatiotemporal activity sequences found from the latent Spatiotemporal features. This HMM identified the hidden dependent relationship between the activity sequences over the Gaussian distributed model. The extracted spatiotemporal features were modeled as Gaussian distribution to predict human mobility. However, the prediction accuracies of this methodology in the afternoon period time are low.

A novel approach [8] was proposed to analyze a surveillance system to obtain human activities. The activities of subjects in the video were initially decomposed into elementary poses and the coefficients of the poses shared by various subjects were denoted by the common coefficients. The activities and the relationships among activities were represented as a graph. The nodes of the graph represented the primary poses and the edges between nodes representing the relationship degree between activates. During the training process, the primary poses and their coefficients were analyzed to obtain possible changes for activities. The probability-based model was used to denote the changes obtained in the activities for each subject. Then the calculated probability model was used to predict the new activities of subjects. However, the graph methods for representing poses and their coefficients might not be sufficient to handle the complex activities.

An abnormal event detection algorithm [9] was proposed by analyzing the information changed over the video sequences. The information from video sequences was represented as an image descriptor. The image descriptor represented the flow of information changes over the video frames in the encoded form. Then a Hidden Markov Model (HMM) was used to represent the state of information from frame to frame with the probability values. The HMM model classified the activities in the video as normal or abnormal activities based on the image descriptor. However, HMM often have a large number of unstructured parameters which affects the performance of abnormal event detection.

A divergence metric based method [10] was proposed for the detection of abnormal activities from crowd. The crowd changes states from frame to frame were found and calculate the amount changes as divergence and curl. The fluid mechanics was used to model the motion changes of crowds between frames. Finally, the changes in the crowd behavior model were measured by using sequential context of motion which detected the behavior changes in the crowd. However, the divergence-curl driven framework was not suitable for lower camera viewpoint.

A novel approach [11] was presented for face recognition based on dynamic rank representation. However, this approach does not applicable for high dimensional data. A Multi-Task Facial Inference Model (MT-FIM)[12] was proposed for face identification and facial expression recognition. Sometimes, MT-FIM may misclassify facial expressions. Adaptive local feature descriptors [13] were proposed for face recognition based on Weber's law. However, it has a low recognition rate. A novel approach based on discriminative learning [14] was presented to recognize humans and their interaction in the video frames. However, using facial descriptor for segregation brought only low successful results. A method [15] was proposed based on Locality Preserving Projection (LPP) and Laplacian faces for face recognition. However, the reconstruction of datas from LPP was very difficult

3. Proposed Methodology

In this section, the proposed method for face detection and face recognition are described in detail. Initially, the abnormal activities of human are predicted by Spatio-Temporal Frequent Object Mining (STFOM) method. Then faces of those detected human are detected by using Viola-Jones face detector and the detected faces are recognized by Orthogonal Locality Preserving Projection (OLPP) where Weighted Local Binary Patterns (WLBP) features and Histogram of Oriented Gradient (HOG) features are used. However, the captured images from videos usually contain significant pose variation. For pose-invariant face recognition, some more features such as Histogram of Face Orientation (HFO) and Histogram of Face Direction (HFD) are included in OLPP based face recognition method.

3.1. Face Detection

The abnormal activities of a person are detected by using STFOM method. The faces of that person are detected by using Viola-Jones algorithm [16]. The Viola-Jones algorithm is used because of its high detection rate. It is more effective in frontal images of faces. Integral image, Ada Boost and the cascade structure are the three main concepts of Viola-Jones allow running in real-time. The integral image is a cost-effective technique for the generation of the sum of the pixel intensities in a specified smaller rectangle in an entire image. It helps the rapid computation of Haar-like features. Computation of the sum of a small rectangular area inside the original image is extremely efficient. It requires only four additions for any arbitrary rectangular size. Ada Boost learning algorithm is used to select the features and assigns a weight to all the selected features. Then, it constructs a strong classifier as a linear combination of weak classifiers. In the cascade structure, combine the classifiers in a cascade. It allows background regions of the image to be quickly discarded. It radically reduces the computation time for face detection.

3.2. Feature Extraction

After the face detection, extract the features in the images using Histogram of Oriented Gradients (HOG) and multi-scale weighted Local Binary Patterns (WLBP) feature descriptors.

3.2.1. Histogram of Oriented Gradients features

HOG is a feature descriptor which counts occurrences of gradient orientation in localized portions of an image. For an input image X, the main steps of HOG are given as follows:

1. Split the image into $n \times n$ non-overlapping parts. Each part contains $c_1 \times c_2$ pixels.
2. Built a block by combining $b_1 \times b_2$ cells. Two adjacent blocks can overlap.
3. For each pixel, $I(x, y)$ the gradient magnitude $m(x, y)$ and orientation $\theta(x, y)$ are computed by

$$dx = I(x+1, y) - I(x-1, y) \tag{1}$$

$$dy = I(x, y+1) - I(x, y-1) \tag{2}$$

$$m(x, y) = \sqrt{dx^2 + dy^2} \tag{3}$$

$$\theta(x, y) = tan^{-1}\left(\frac{dx}{dy}\right) \tag{4}$$

4. Split the orientation range into k bins and then compute the histogram within part (HC)

$$HC(k)_i = HC(k)_i + m(x,y), \text{if } I(x,y) \in cell_i \text{ and } \theta(x,y) \in bin(k) \quad (5)$$

5. The histogram of a block HB can be obtained by integrating the HCs within this block possible.

$$HB_j = \{HC_1, HC_2, \dots HC_{b1\times b2}\} \quad (6)$$

6. Normalize the vector of $HB_j(NHB_j)$ by $L2$-norm block normalization

$$(NHB_j) = \frac{HB_j}{\sqrt{\|HB_j\|_2^2 + e^2}} \quad (7)$$

where, e is a constant used to avoid the problem of division by zero. If there are N blocks in an image, the final histogram can be obtained by integrating all normalized block's histograms.

$$HOG = \{NHB_1, NHB_2, \dots, NHB_j, \dots NHB_N\} \quad (8)$$

3.2.2 Multi-scale weighted Local Binary Patterns Features

The standard version of LBP of a pixel is formed by thresholding the 3×3 neighborhood of each pixel value with the center pixel's value. Assume, g_c be the center pixel gray level and g_i be the gray level of each surrounding pixel. Let $LBP_{p,r}$ be the LBP feature of a pixel's circularly neighborhoods, where r is the radius and p is the number of neighborhood points on the circle.

$$LBP_{p,r} = |s(g_{p-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{p-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \quad (9)$$

where $s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & otherwise \end{cases}$.

In the LBP, the similar patterns are positioned in one bin of the histogram, so the dissimilar patterns are lost. To sort out this problem, a multi-scale weighted LBP is introduced. In WLBP, LBP features with pretty large radius is extracted and for those pixels whose LBP values are non-uniform patterns, the LBP feature with smaller radius are further extracted. For those pixels whose new LBP values are relatively dissimilar patterns, the LBP-features with even smaller radius are further extracted. This process is repeated until the LBP features with minimum radius are extracted. This method centers on a weighted distance function for two images. Each image is represented by an LBP vector by using LBP operator at a given scale. Here, a scale is a natural number that is used to multiply the radius values utilized by the selected LBP operator to produce the actual radius values used in the computation. Consider, m and m' be two images, $V_1, V_2, \dots, V_n$ be the LBP feature vectors for image m extracted by LBP operator at scales $\{1,2,\dots,n\}$, $V_1', V_2', \dots, V_n'$ be the LBP feature vectors for image m' and $w_1, w_2, \dots w_n$ be the weights of the scales. The distance between two images is defined as follows:

$$d(m,m') = w_1 \times \|V_1 - V_1'\| + w_2 \times \|V_2 - V_2'\| + \dots + w_n \times \|V_n - V_n'\| \quad (10)$$

are where, $\|V_i - V_i'\|$ is the Euclidean distance of vectors V_i and V_i'. In the WLBP features, there are multiple radius parameters to be scaled.

The extracted HOG feature represents the edge structure of theface and the WLBP feature can capture the local texture information about a face.

3.3 Face Recognition

Face recognition is used to verify individual identity. Here, Orthogonal Locality Preserving Projection (OLPP) is used to recognize the faces. The overall process of OLPP based face recognition is given as follows:

1. **Principal Component Analysis (PCA) Projection**: The components related to zero eigenvalue are thrown away and the face images x_i is projected into the PCA subspace. W_{PCA} denotes the transformation matrix of PCA. By PCA projection, the extracted HOG and WLBP features are statistically uncorrelated. The rank of the new data matrix is equal to the number of features.

2. **Adjacency Graph Construction:** Let G be a graph with n nodes. The i-th node corresponds to the face image x_i. An edge is drawn between the nodes i and j when the x_i and x_j are close. That is, if x_i is among p nearest neighbors of x_j or x_j is among p nearest neighbors of x_i, an edge is drawn between i and j.

3. **Weights Selection:** If node i and j are connected, assign

$$S_{ij} = e^{-\frac{\left\|x_i - x_j\right\|^2}{t}} \tag{11}$$

Otherwise, assign$S_{ij} = 0$. The local structure of the face is modeled by the weight matrix S of graph G.

4. **Calculation of Orthogonal Basis Functions:** A diagonalmatrix is represented as D whose entries are sums of S, $D_{ii} = \sum_j S_{ji}$. Laplacian matrix $L = D - S$ is defined in spectral graph theory. Let $\{a_1, a_2, \dots, a_k\}$ be the orthogonal basis vectors, it is defined as follows

$$A^{(k-1)} = [a_1, a_2, \dots, a_{k-1}] \tag{12}$$

$$B^{(k-1)} = \left[A^{(k-1)}\right]^T (XDX^T)^{-1} A^{(k-1)} \tag{13}$$

The orthogonal basis vectors $\{a_1, a_2, \dots, a_k\}$ can be calculated as follows.

Calculate a_1 as the eigenvector of $(XDX^T)^{-1}XLX^T$ associated with the smallest eigenvalues where $X = [x_1, x_2, \dots x_n]$ be a set of face images.

Calculate a_k as the eigenvector of

$$M^{(k)} = \left\{I - (XDX^T)^{-1} A^{(k-1)} \left[B^{(k-1)}\right]^{-1} \left[A^{(k-1)}\right]^T\right\} . (XDX^T)^{-1} XLX^T \tag{14}$$

associated with the smallest eigen value of $M^{(k)}$.

5. **OLPP Embedding:** Let $W_{OLPP} = [a_1, \dots a_l]$ the embedding is as follows:

$$x \to y = W^T x \tag{15}$$

$$W = W_{PCA} W_{OLPP} \tag{16}$$

where W is the transformation matrix and y denotes l-dimensional representation of the face image x.

6. **Locality preserving function:** The OLPP tries to preserve the local geometric structure of the face images. OLPP finds the basis vectors by minimizing the locality preserving function which is given as follows:

$$f(a) = \frac{a^T XLX^T a}{a^T XD \;\; ^T a} \tag{17}$$

$f(a)$ specifies the locality preserving the power of the projective map a.

3.4 Pose Invariant Face Recognition

The face images may be detected in the unconstrained environment which is usually containing pose variation. The OLPP with HOG and WLBP features is designed to recognize the frontal faces. To recognize the faces in different poses, pose invariant face recognition is introduced. In the pose invariant face recognition, the detected face images were split into $M \times N$ overlapped patches. The severity of occlusion for each patch is then calculated based on the detected boundary between the occluded and unoccluded facial texture. Then check whether 80% of pixels in one patch fall into the unoccluded region. If it is so, then it is chosen as an unoccluded patch else the patch is ignored due to the large area of occlusion. After that, each of the unoccluded patches is split into $J \times J$ cells. Then along with the HOG and WLBP features, Histogram of Face Orientation (HFO) and Histogram of Face Directions (HFD) are included in the OLPP face recognition method.

The HFO feature is a feature which is calculated based on the count of face orientations for each angle in an image. That is, to say, the HFO feature is the distribution of face orientation frequencies in an image. This HFO feature descriptor has 13 feature dimensions which correspond to 15° resolution in $[-90^{\circ}, 90^{\circ}]$ interval. The HFD feature comprises the distribution of direction frequencies in the images. It is formed from a coarser histogram representation of the face orientations. Directions are defined as left, front and right. These correspond to the angle intervals$[-90^{\circ}, 45^{\circ}]$, $[-30^{\circ}, 30^{\circ}]$ and $[45^{\circ}, 90^{\circ}]$ respectively. The HFD feature descriptor is coarser for of HFO and it has three dimensions. These features are used in the OLPP for face recognition. It effectively recognizes the faces event at different poses.

4. Result and Discussion

In this section, the efficiency of Spatio-Temporal Frequent Object Mining- Viola-Jones based Face Detection-Orthogonal Locality Preserving Projection (STFOM-FD-OLPP), STFOM-FD- Pose Invariant OLPP (STFOM-FD-PIOLPP) and STFOM-FD-LPP [15] is tested in terms of accuracy, information gain ratio and true positive rate. In STFOM-FD-LPP, LPP method is applied for face recognition. A dataset in [3] is used in this experiment to analyse the efficiency of STFOM-FD-OLPP and STFOM-FD-PIOLPP.

4.1 Accuracy

Accuracy is calculated as the ratio of the relevant and total number of detected and recognized human faces who involved in the abnormal activity. It is calculated by dividing the number of correctly detected and recognized human faces who involved in the abnormal activity by the total number of instances. It is given as follows:

$$Accuracy = \frac{True\ Positive\ +\ True\ Negative}{True\ Positive\ +\ True\ Negative + False\ Positive +\ False\ Negative}$$

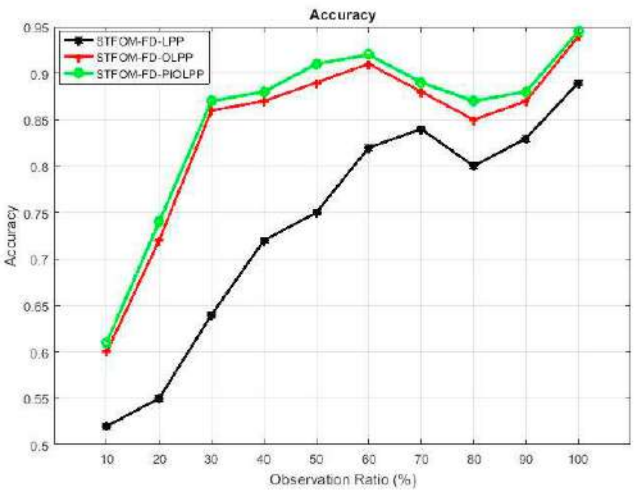

Fig. 1. Comparison of Accuracy

Fig. 1 shows the comparison of accuracy between STFOM-FD-OLPP and STFOM-FD-PIOLPP methods for different observation ratio. The observation ratio is taken in X-axis and the accuracy is taken in Y-axis. When the observation ratio is 50%, the accuracy of STFOM-FD-PIOLPP is 21.33% greater than STFOM-FD-LPP, 2.24% is greater than STFOM-FD-OLPP. By considering orientation and direction based features in STFOM-FD-PIOLPP method, the faces are detected and recognized more accurately than STFOM-FD-OLPP and STFOM-FD-LPP methods. From this analysis, it is proved that the proposed STFOM-FD-PIOLPP has better accuracy than the other STFOM-FD-OLPP method.

4.2 Information Gain Ratio

Information Gain Ratio is defined as the amount of information gain while detecting and recognizing the faces of human in the videos.

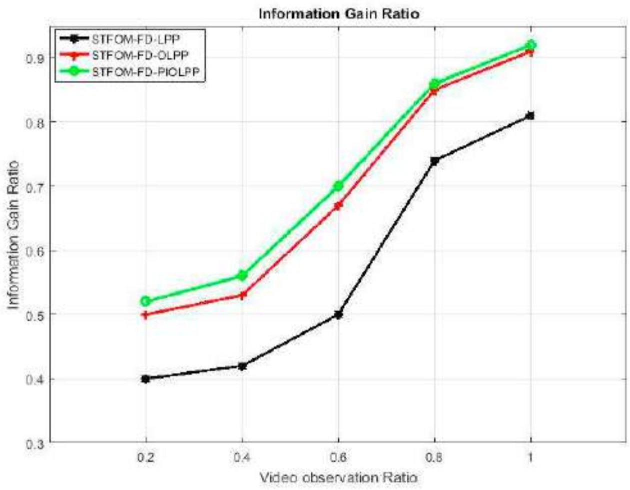

Fig. 2. Comparison of Information Gain Ratio

Fig. 2 shows the comparison of information gain ratio between STFOM-FD-OLPP and STFOM-FD-PIOLPP methods for different video observation ratio. The video observation ratio is represented in X-axis and the information gain ratio is represented in Y-axis. When the observation ratio is 0.2, the information gain ratio of STFOM-FD-PIOLPP is 30% greater than STFOM-FD-LPP and 4% is greater than STFOM-FD-OLPP. By using HOG, WLBP, HFO and HFD features in OLPP, the information gain ratio of proposed STFOM-FD-PIOLPP is higher than STFOM-FD-LPP and STFOM-FD-OLPP methods. From this analysis, it is proved that the proposed STFOM-FD-PIOLPP has better information gain ratio than the other STFOM-FD-OLPP method.

4.3 True Positive Rate

True Positive Rate is defined as the rate of suspicious human faces is recognized as a suspicious human in videos.

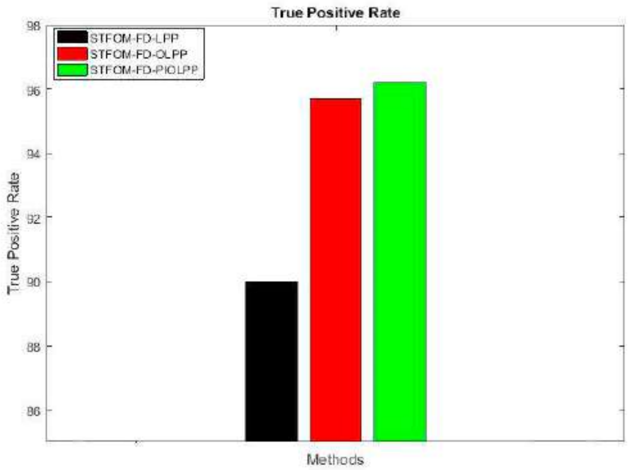

Fig. 3. Comparison of True Positive Rate

Fig. 3 shows the comparison of true positive rate between STFOM-FD-OLPP and STFOM-FD-PIOLPP methods. The methods are represented in X-axis and the true positive rate is represented in Y-axis. The true positive rate of STFOM-FD-PIOLPP is 6.8% is greater than STFOM-FD-LPP and 0.52% is greater than STFOM-FD-OLPP. The reason behind the high true positive rate of STFOM-FD-PIOLPP is that it considered the pose variations which induce the look change in the face image. This considerably increases the true positive rate of STFOM-FD-PIOLPP for face recognition. From this analysis, it is proved that the proposed STFOM-FD-PIOLPP has better true positive rate than the other STFOM-FD-OLPP method.

5. Conclusion

The accurate detection and recognition of the human faces for confirmation of abnormal human activity is the main objective of this paper. The pose invariant based face recognition is also the main contribution of this paper. Initially, STFOM mined the abnormal activities of human in the video surveillance system. Then, a Viola-Jones face detector is applied to detect the faces of human. The HOG and WLBP features are extracted from the detected faces which are used in the OLPP to recognize the faces. In the procedure of OLPP, PCA is used to project the face images into subspaces. However, the detected faces have pose variations the HOG and WLBP feature is more effective for frontal faces. So pose invariant based OLPP face recognition is introduced for face recognition where the HFO and HFD are used to define the orientation and direction of faces. By using HFO and HFD features for face recognition, the accuracy, information gain ratio and true positive rate of STFOM-FD-PIOLPP are better than STFOM-FD-OLPP and STFOM-FD-LPP method.

References

[1] Taha, Ahmed, Hala H. Zayed, M. E. Khalifa, and El-Sayed M. El-Horbaty. (2015) "Human activity recognition for surveillance applications." in Proceedings of the *7th International Conference on Information Technology* . 577-586.

[2] Li, Kang, and Yun Fu. (2014) "Prediction of human activity by discovering temporal sequence patterns." *IEEE transactions on pattern analysis and machine intelligence***36**(**8**): 1644-1657.

[3] Manju, Doraisami, and Radha, V. (2018) "An Efficient Spatiotemporal Frequent Object Mining Method to Predict Abnromal activities."*Journal of Advanced Research in Dynamical and Control Systems (JARDCS)***10**(**3**): 1025-1029.

[4] Sharif, Md Haidar, and ChabaneDjeraba. (2012) "An entropy approach for abnormal activities detection in video streams." *Pattern Recognition***45**(**7**):2543-2561.

[5] Forkan, Abdur Rahim Mohammad, Ibrahim Khalil, Zahir Tari, SebtiFoufou, and Abdelaziz Boura. (2015) "A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living." *Pattern Recognition***48**(**3**): 628-641.

[6] Singh, Dinesh, and C. Krishna Mohan. (2017) “Graph formulation of video activities for abnormal activity recognition.” *Pattern Recognition***65**: 265-272.
[7] Guo, QiuLei, and Hassan A. Karimi. (2017) “A novel methodology for prediction of spatial-temporal activities using latent features.” *Computers, Environment and Urban Systems***62** (**2017**): 74-85.
[8] Augustyniak, Piotr, and GrażynaŚlusarczyk. (2018) “Graph-based representation of behavior in detection and prediction of daily living activities.”*Computers in biology and medicine***95**(**2018**): 261-270.
[9] Wang, Tian, MeinaQiao, Yingjun Deng, Yi Zhou, Huan Wang, Qi Lyu, and HichemSnoussi. (2018) “Abnormal event detection based on analysis of movement information of video sequence.” *Optik***152***:* 50-60.
[10] Chen, Xiao-Han, and Jian-Huang Lai. (2019) “Detecting abnormal crowd behaviors based on the div-curl characteristics of flow fields.”*Pattern Recognition***88**: 342-355.
[11] Li, Hongjun, and Suen, Ching. (2016)“Robust face recognition based on dynamic rank representation.”*Pattern Recognition***60**: 13-24.
[12] Zheng, Hao, Geng, Xin, Tao, Dacheng, andJin, Zhong. (2016) “A multi-task model for simultaneous face identification and facial expression recognition.”*Neurocomputing***171**: 515-523.
[13] Yang, Wankou, Wang, Zhenyu, and Zhang, Baochang. (2016) “Face recognition using adaptive local ternary patterns method.”*Neurocomputing***213:** 183-190.
[14] Tanisik, Gokhan, Zalluhoglu, Cemil, andIkizler-Cinbis, Nazil. (2016) “Facial descriptors for human interaction recognition in still images.”*Pattern Recognition Letters***73**: 44-51.
[15] Bansal, Prachi, Mittal, Saurabh, and Gupta, Maanak. (2014) “Using locality preserving projections in facial recognition.”*International Journal of Engineering Technology, Management and Applied Sciences***2**(**3**): 99-108.
[16] Viola, Paul, and Michael J. Jones. (2004) “Robust real-time face detection.”*International journal of computer vision***57**(**2**): 137-154.