**Australian Journal of Basic and Applied Sciences**, 9(23) July 2015, Pages: 418-423

ISSN:1991-8178

**Australian Journal of Basic and Applied Sciences**

Journal home page: www.ajbasweb.com

# Categorization of Video Using Viola Jones and Fisher's Linear Discriminant Function

[1]Shoba Rani and [2]Syed Ali. D.

*[1]Asso.Prof, [2]Asst.Prof, Dr. M.G.R. Educational and Research Institute University*

**ARTICLE INFO**

***Article history:***
*Received 12 March 2015*
*Accepted 28 June 2015*
*Available online 22 July 2015*

***Keywords:***
*Video frames, Viola Jones algorithm, Object Tracking Algorithm, Fisher's Linear Discriminant function and Cascading Classifiers*

**ABSTRACT**

This paper presents the categorization of objects and human beings. For categorization process frame by frame search is made on videos in a video pool. Frames are extracted from Video using Viola Jones Algorithm and categorized by using Fisher's Linear Discriminant Function. The results of categorization is good.



**To Cite This Article:** Shoba Rani and Syed Ali. D., Categorization of Video Using Viola Jones and Fisher's Linear Discriminant Function. ***Aust. J. Basic & Appl. Sci.,*** *9(23): 418-423, 2015*

## INTRODUCTION

In our today life so many videos are captured using camcorders. Many video clips are available in the internet also. It is difficult to go through all video clips to find an object or person. Categorizing the video contents is not an easy task. When working with static images, a large amount of area must be scanned to detect the contents. In real time applications such as security, video surveillance, biomedical applications and many other fields there is a great need of detecting faces or objects. Human faces are possible to detect because they have common facial features like nose, eyes, mouth. But objects are not having common features. So detection of face in a video takes more importance. Except human faces other contents are considered as objects.

The main approaches are Feature Invariant approach which try to detect features like skin color, shape that are invariant to some conditions like light, pose. Second method is Template matching approach where matching is made between template corresponding to trained images and input images. Third approach is Knowledge-based approach where based on knowledge of human some rules are programmed to detect face. Finally Appearance based approach uses machine learning technique. Some features of face are pre-defined in training stage. Input image is compared with this pre-labeled data set. Some face detection algorithms under this category are Support Vector Machine, neural networks, AdaBoost Based face detection method. The proposed paper uses appearance based approach which gives more accurate result compared to other methods.

Viola Jones method is proposed by Viola P. and Jones, M J (2001). For this method a single video clip is given as an input which contains many human faces and many objects. The faces are detected using pre-defined data and objects are rejected by this method. Rejected objects are collected for other purpose like object detection.

Paul Viola and Michael Jones proposed Robust Real-time Object Detection in 2001. The main contribution was introduction of new image representation called "integral images". They also explained about learning algorithm, based on AdaBoost. Lastly they contributed a method ofcascading the classifiers. They found that faces detection is more accurate compared to other methods.

Daniel Hefenbrock, Jason Oberg, Nhat Tan Nguyen Thanh, Ryan Kastner, Scott B. Baden proposed Accelerating Viola-Jones Face Detection to FPGA-Level using GPUs. They implemented Viola Jones face detection method on GPU's and found that high performance. Theo Ephraim and Tristan Himmelman presented a paper on Optimizing Viola-Jones Face Detection For Use In Webcams. They tried to modify the Viola Jones face detection algorithm for real time implementation. They found that its very slow to implement for webcams.

Ijaz Khan, Hadi Abdullah and Mohd Shamian Bin Zainal in 2012 proposed Efficient eyes and mouth detection algorithm using combination of viola jones and skin color pixel detection. In this method combination of Viola Jones algorithm which

**Corresponding Author:** Shoba Rani, Asso. Prof, Dr. M.G.R. Educational and Research Institute University
E-mail: shobarani.cse@drmgrdu.ac.in

accurately detect faces and skin color pixel detection which takes less time was used. They explained about mouth and eye detection by combining two algorithms. Experiments resulted more accurate detection faces and their features like mouth and eyes. Ramana Isukapalli, Russell Greiner, and Ahmed Elgammal explained Learning a Dynamic Classification Method to Detect Faces and Identify Facial Expression. They discussed about building a decision tree of cascade classifiers. Most effective classifier is cascaded to detect face and expression of face. They used DTC classifiers and found more accurate result in their work.

Pantic, 2000, identified three basic problems in facial expression analysis approach. It deals with face detection in a facial image/image sequence, data extraction from facial expression, and classification of facial expression. Viola, 2001, discussed the Adaboost algorithm was fully permit a search sub window for rapid face detection of the image at multiple scales. Essa, 1997, the spatial and temporal filtering together with thresholding is used to extract motion blobs from image sequences. Turk, 1991, the blobs are then evaluated using the eigen faces method to detect the presence of a face, Principal component analysis (PCA) is the method to calculate the distance of the experiential region from a face space of 128 sample images. Steffens, 1998 the Person Spotter system discussed the spatio temporal filtering and stereo disparity in pixels. It tracks the bounding box of the head, selecting image regions of interest in the video frame also works with skin color and convex region detectors to check for face presence in these regions.

Littlewort, 2002, presented a bank of 40 Gabor wavelet filter to achieve convolution for data extraction at different scales and orientations. Lades 1993, extracted a jet of magnitudes of complex valued responses at different locations in a lattice executed on an image. Essa, 1997, extended their face detection approach to extract the positions of prominent facial features using Eigen features. With the help of FFT, PCA calculates the distance of an image from a feature space.

## *II. Proposed Work:*

Many videos are kept in a video pool and merged into a single video. Categorization is done based on transition clues like objects or human beings. For categorization process frame-by-frame search is made on videos in a video pool. Frames are extracted from Video using Viola Jones algorithm. Features are extracted from the frame using Fisher's Linear Discriminant function. In each frame, a complete object is extracted. The features are extracted from a frame is considered as a pattern. If 25 frames belonging to a video are considered, then 25 patterns are created. This proposed system is mainly used for separating human beings and objects.

So many algorithms are there to detect faces in static images. In this proposed system single-shot, short personal videos are considered. From all these videos frames are extracted for each movement of camcorder and content of video. The proposed system mainly takes two stages for categorizing a video, preprocessing stage and categorization stage.

### *A. Preprocessing stage:*

This stage takes many short videos of different resolution and extracts frame from all videos. Each pixel value in a frame compared with the threshold value for removing non informative pixels. After preprocessing stage object and human frames are extracted. All short videos are merged into a single video of static resolution.

### *Track to locate face:*

Face tracking can be analyzes video frames and find the position of moving faces within the each frame can be done by the algorithm. For each track to locate face, three steps are followed: Initialization, tracking, stopping. Many methods used a face detector for initialization of their tracking processes. There have been studies on the profile pose face detectors. They suffered from the false-detection problem. Chaudhury et al., 2003, discussed that instead of a fixed threshold value to initialize the face tracker, used two face probability maps are frontal views and profiles. The face probabilities are transmitted throughout the temporal sequence. The non-faces are classified and eliminated by probabilities. They either go to zero or remain low. To represent an intermediate head pose, information from the two faces probability maps are combined. The experiments revealed that the probabilistic detector was very accurate than a traditional face detector. Boccignone et al., 2005 and Li et al., 2006, described that certain type of processed video, which depends on, a learning set dedicated to a color based face trackers. It might not work on unknown videos with changing illumination conditions. They are more robust to various illuminations and occlusions. Arnaud et al., 2005; Zhu et al., 2005; Tong et al., 2007, explained the facial features extraction from eyes, nose and mouth. Facial features always support tracking of high level facial information.

In general, the following procedure is approved for the object tracking algorithm.

**Step 1:** Wait for object/objects to appear in the frame
**Step 2:** Enter initialization mode (wait for a object/objects to appear for a predefined amount of time).
**Step 3:** Enter tracking mode to choose the closest object.
**Step 4:** Until it leaves the frame, track the object. Leave tracking mode, when the tracked object disappears, for a predefined amount of time to avoid losing track of the object due to minor movements.
**Step 5:** Go to **step 1**.

***B. Categorization stage:***

In this stage human and object are separated. It takes extracted frames which have most of the information's. These are subjected to Viola Jones algorithm. This algorithm is mainly used for detecting faces. Viola-Jones have devised a scale invariant detector that requires the same number of calculations whatever the size.

The algorithm must be trained before detection process. For training process two samples are given. One is positive sample which contain only faces. Second is negative samples which are object other than human faces. A sub window is drawn on an image many times to detect features of face.

***B1. Scale Invariant Object Detection:***

Viola-Jones has devised a scale invariant detector that requires the same number of calculations whatever the size. In this stage the input frames are converted into integral images. In Integral image, each pixel is equal to sum of all pixels above and to the left of the concerned pixel.

1: Input: an image I of size N × M.
2: Output: its integral image II of the same size.
3: Set II(1, 1) = I(1, 1).
4: for i = 1 to N do
5: for j = 1 to M do
6: II(i, j) = I(i, j) + II(i, j − 1) + II(i − 1, j) − II(i − 1, j − 1) and II is defined to be zero whenever its argument (i, j) ventures out of I's domain.
7: end for
8: end for

Input image pixel

| | | |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 1 | 1 |
| 1 | 1 | 1 |

Integral image pixel

| | | |
|---|---|---|
| 1 | 2 | 3 |
| 2 | 4 | 6 |
| 3 | 6 | 9 |

Sum of all pixel values in a given rectangle can be calculated by taking four points in a rectangle at corners.

***B2. Human being or object feature representation using Fisher's Linear Discriminant Function:***

Object feature extraction is the process of converting a object image into a feature vectors. The feature vector carrying characteristics of the object after the detection of object, holistic Approach or analytic Approach is used to extract the features: In holistic methods, raw facial image is exposed for feature extraction. Though, in analytic, some important facial features are detected. Here, we use Holistic Approach, as that it means we send a raw image. There are three types of feature extraction techniques are used. Namely, Feature-based method, appearance based method and hybrid method. The geometric features like facial points or shapes of facial components or spatial locations are used by feature-based methods for e.g. FACS. The colors, colors layouts or textures of the facial skin including wrinkles and furrows are used by the appearance-based method. Namely Local feature based method. The geometric and appearance facial features are used by hybrid methods. The facial emotional features are extracted either from the entire image or from regions. The whole image feature is simpler. While region based representation of emotional images is proved to be more accurate.

The steps involved in the linear mappings are

**Step 1:** Computation of the Discriminant vectors φ1 and φ2. This is exact for a particular linear mapping algorithm.

**Step 2:** Computation of the planar images of the original data points: this is common for all linear mapping algorithms.

***B3. Cascading Classifiers:***

In this method strong classifiers are constructed by the combination of weak classifiers. It is machine learning methods which detects best features. Many strong classifiers are cascaded to detect face. The output of each classifier is fed as input to next classifier. It will discard non face region easily and detect human faces. Haar wavelet transformation: It takes frames and compress to threshold. This helps in less computation.

***III. Implementation:***

The implementation detail for categorization is show as steps below:

Step1: Take different resolution videos as input.
Step 2: Remove non informative pixels by comparing with threshold value.
Step 3: Convert all videos into static resolution.
Step 4: Extract frames and merge frames.
Step 5: Merged frames are converted to video.
Step 6: single video is compared with training data for face detection.
Step 7: If facial features detected move frames to HUMAN FACE folder. Else move frames to OBJECT folder.
Step 8: Repeat step 6 until the last frame of video.

For preprocessing many videos short, single-shot videos are given as input. Videos are merged to form a single video.

Single video is given as input to Viola Jones algorithm. Detected Faces are marked by rectangular window. Objects are rejected and collected in one folder.

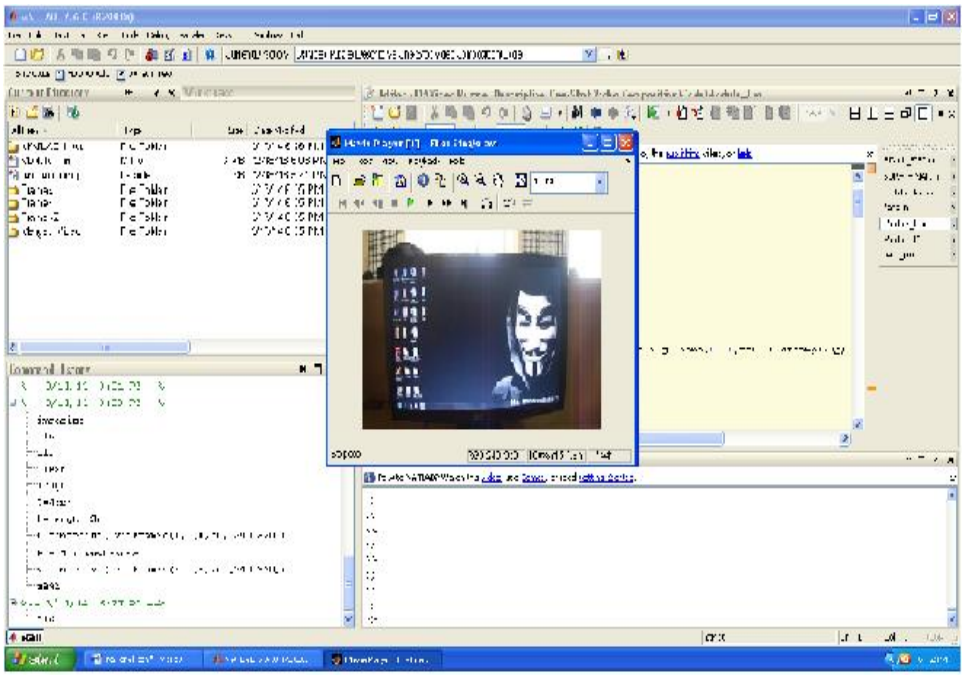

**Fig. 1:** Merged video.

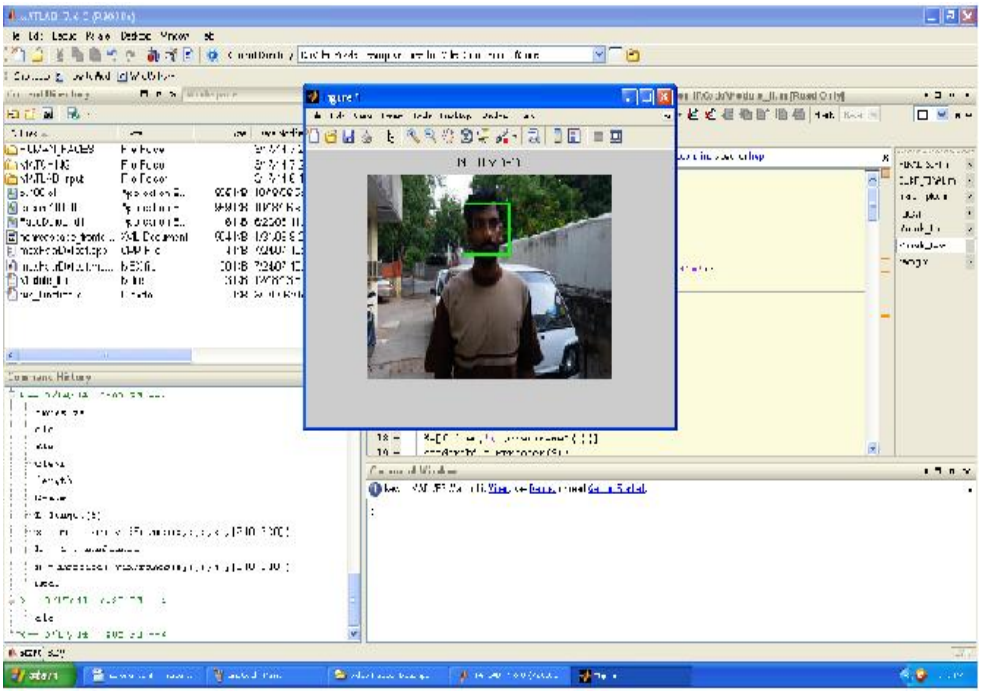

**Fig. 2:** Face detection.

After algorithm finishes scanning all frames of a video two folders are created one is MATCHING folder containing object frames. Other is HUMAN_FACES folder containing detected human faces frames.

This proposed work is successfully separating the human beings and objects. This shows accurate categorization of video.

### *IV. Conclusion and Future Work:*

Many single-shot, short videos can be merged to form a video and content of that video is categorized accurately as human frames and object frames. The future work can be applied to videos taken in natural environment for different types of expressions simulated under various possible conditions like responses to expectations from other people.

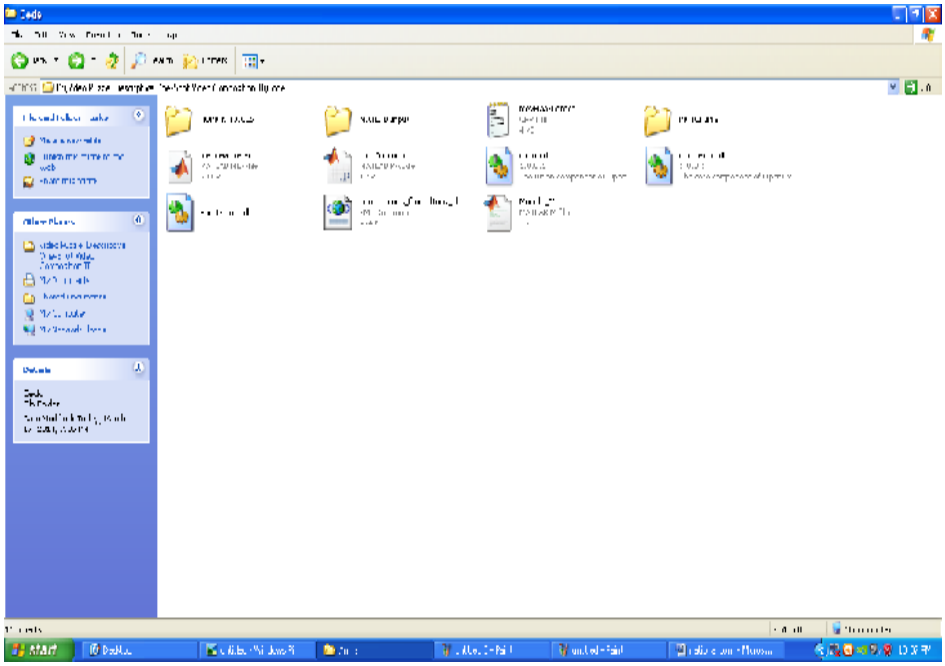

**Fig. 3:** Folders after categorization.

**Fig. 4:** Human faces collected in one folder.

## REFERENCES

Rong Xiao, Ming-Jing Li, Hong-Jiang Zhang, 2004. Robust Multi-Pose Face Detection in Images, http://research.microsoft.com/pubs/69036/csvt_revision.pdf, 14(1): 31-41.

Ijaz Khan, Hadi Abdullah and Mohd Shamian Bin Zainal, 2013. Efficient Eyes and Mouth Detection Algorithm using Combination of Viola Jones and Skin Color Pixel Detection, http://www.eaas-journal.org/survey/userfiles/files/v3i407%20CS%20Engineering.pdf, 3: 51-60.

Daniel Hefenbrock, Jason Oberg, Nhat Tan Nguyen Thanh, Ryan Kastner, Scott B. Baden, 2010. Accelerating Viola-Jones Face Detection to FPGA-Level using GPUs, http://bamboo.ucsd.edu/tan/publish/fccm_10.pdf.

Ramana Isukapalli, Russell Greiner and Ahmed Elgammal, 2005. Learning a Dynamic Classification Method to Detect Faces and Identify Facial Expression,

http://www.cs.rutgers.edu/~elgammal/pub/isukapalli_AMFG05_FaceExpr.pdf., 3723: 70-84.

Anuj Bhardwaj and Rashid Ali, 2009. Image Compression Using Modified Fast Haar Wavelet Transform, http://www.idosi.org/wasj/wasj7%285%29/13.pdf.

Munther Abualkibash, Ahmed Elsayed, Ausif Mahmood, 2013. Highly Scalable, Parallel and Distributed Adaboost Algorithm using Light Weight Threads and Web Services on A Network of Multi-Core Machines, http://arxiv.org/ftp/arxiv/papers/1306/1306.1467.pdf.

Pantic, M. and L.J.M. Rothkrantz, 2000. Expert system for Automatic Analysis of facial Expressions, ELSEVIER, Image and Vision Computing, 18(11): 881-905.

Viola, P. and M. Jones, 2001. Robust real-time object detection, Technical Report 2001/01, Compaq Cambridge Research Lab.

Essa, I.A. and A.P. Pentland, 1997. Coding, analysis, interpretation and recognition of facial expressions, IEEE Transaction on Pattern Analysis and Machine Intelligence, 19(7): 757-763.

Turk, M. and A. Pentland, 1991. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3(1): 71-86.

Steffens, J., E. Elagin and H. Neven, 1998. Person spotter fast and robust system for human detection, Tracking, and recognition, In Proceedings International Conference on Automatic Face and Gesture Recognition, pp: 516-521.

Littlewort, G., I. Fasel, M. Stewart, Bartlett and J.R. Movellan, 2002. Fully automatic coding of basic expressions from video, Technical Report, UCSD INC MP Lab.

Boccignone, G., V. Caggiano, G.D. Fiore and A. Marcelli, 2005. Probabilistic detection and tracking of faces in video, In Proceedings of International Conference on Image Analysis and Processing, pp: 687-694.

Arnaud, E., B. Fauvet, E. Memin and P. Bouthemy, 2005. A robust and automatic face tracker dedicated to broadcast videos, In Proceeding Of International Conference on Image Processing, pp: 429-432.

Li Y., H. Ai, C. Huang and S. Lao, 2006. Robust head tracking with particles based on multiple cues fusion, ECCV Workshop on HCI, pp: 29-39.

Lades, M., J.C. Vorbruggen, J. Buhmann, J. Lange, C. vonder Malsburg, R.P. Wurtz and W. Konen, 1993. Distortion invariant object recognition in the dynamic link architecture, IEEE Transactions on Computers, 42(3): 300-311.