IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 8 (August. 2019), ||S (I) || PP 26-39

Critical Analysis of Various Approaches for Automatic Human Face Image Retrieval

Faizur Rashid[1], Zafar G. Sheikh[2]
[1]*Pragnya College of Management and Computer Studies, Pune, India*
[2]*Department of Information System, College of Computing and Informatics, Haramaya University*

Abstract: With enormous growth of security and surveillance system, a huge amount of video data is being generated every day. It is immense challenge for researcher to search and retrieve human face of interest from video. The proposed work is inspired from the same issue in concern with analyzed by approach and application. It would be the future demand for searching, browsing, and retrieving human face of interest from video database for several applications.
To understand and satisfy the future demands, proper analysis and evolution of approach is the necessity of time. The objective of this paper is to analyze the available human face retrieval approaches and methods in various media files. The various media files play an important role in everyday life such as feature-length, news video, state-of-art application, movie or any other video and to propose more acceptable methodology. The works fulfilled under the three objectives-i) classification of human face retrieval in different media files with application. ii) Techno-critical exploration of different human face retrieval approaches from media files which analyzed the technical details. Finally, iii) detail investigation to leads fulfilled the different objectives in avail the application.

Keywords: face retrieval, face recognition, face detection, face tracking

Date of Submission: 29-07-2019 Date of Acceptance: 14-08-2019

I. INTRODUCTION

The proposed work is inspired from the issue to search and retrieve human face of interest from video. It would be the future demand for searching, browsing, and retrieving human face of interest from video database for several applications such as state-of art, security, and surveillance, personal and industrial demands. Human face retrieval is the collective work of major aspects such as detection, tracking, recognition. It is also called the process. The researcher has been successful demonstrated the implementation of detection and recognition. Still, the real life problems are yet to solve.

The proposed work has focuses on video database, as per our analysis, it has not acquire much attention and success as compare to detection and recognition because of two reasons-i) necessity is the mother of invention, i. e. researcher still not consider an important issue. ii) It is mutual facility which includes various approaches to satisfy the aim and objective which changes with time. The paper aims to analyze and study the available human face retrieval approaches and methods for the betterment of specified application. The various media files play an important role in everyday life such as feature-length, news video, state-of-art application, movie or any other video. The presented works aims to fulfill the three objectives-i) classification of human face retrieval in different media files with application. ii) Techno-critical exploration of different human face retrieval approaches which analyzed and evaluated the technical details using various stages. Finally, iii) detail investigation to leads fulfilled the different objective avail the application. The paper, categorize human face retrieval into four basic categories -Holistic based (3-D model), Part-based (tracker), Color based and Multi cue (fusion) based.

Human face detection and recognition from video database is very intuitive to computer and human, still various challenges to computer for face detection described in [1] like pose, scale, illumination, expression etc. In concert with video, work is in progress to overcome the problems in complex background [2] by detecting and tracking video images. Whereas, skin-tone has fine prospect provided for illumination invariant [3]. Face detection under partial occlusion [4] using components and their topology. Face detection in video sequences [5] and localization [6]. Face detection and tracking performed in [7]. As a part of face retrieval, recognition has important aspect. To simplify the issues related with recognition work is in progress in diverse situation such as, color face recognition [8], large scale face recognition in social network [9, 12]. The global and local information used in [10] and Isogeodesic Stripes used for 3-D face recognition [11]. However, Maximum Correntropy Criterion used for Robust Face Recognition [13]

The paper is organized as follows- Section II described classification of face retrieval in various media files, the categorization according to appearance in Section III, and Section IV important tabular analysis of different approaches, the performance evolution and results in Section V. The paper ends with conclusion in Section VI.

II. CLASSIFICATION IN VARIOUS MEDIA FILES

The various media files play an important role in everyday life such as feature-length, news video, state-of-art application, movie or any other video. Media have different feature such as visual, audio or textual/script.

Media → / Features ↓	Film/ movies	Video/ TV	Real-time	Other
Visual	• Character retrieval • Character identification • People finding	• Person spotting • Person identification • Indentifying individuals • Face sequence matching • Face track finding • Face image retrieval • Face retrieval	• Person identification	• Face region detection • Actor based indexing
Audio-visual	• Speaker retrieval	• Person retrieval • Major Cast detection		
Script-Visual	• Character identification	• Naming of characters • Names and faces		

Table 1. Classification of face retrieval in various media files

Face retrieval approaches proposes the one or more combination of available features to assure the endeavor from different media. Table 1 shows the different endeavors on media along with existing features. The classification is performed according to available media such as film/movies, video/TV, real time and other domain like compressed. The features has classified into three major classes according to the application used for the retrieval process. First, visual features which considered only the facial features from image/frame. It includes holistic and part-based approaches as major. The second classification is audio-visual features, the voice segmentation and detection fused with visual features for better identification results. The third classification with script-visual, the movies script or text is involved for the person identification/ verification. The script/ text are segmentation and processes with visual features. The second and third classification includes multi-clue based approaches.

III. CATEGORIZATION OF FACE RETRIEVAL

In this paper the approaches of face retrieval has been categorized into four groups, from the common approach or methodology used to achieve the goal.

A. Holistic(3-D MODEL) BASED

The holistic approach generally refers to methods that use the entire face image for face identification. Basically its includes methods like Principal Component Analysis (PCA) or Eigenfaces, Linear Discriminate Analysis (LDA) or Fisherface. We consider the recent work using 3-D modeling of face and head.

1 . Automated person identification in video

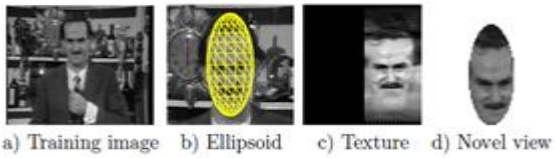


Fig 1. Ellipsoid head model

In [15], authors proposed method to identify the target face with variation in pose, illumination and face expression. First step is to detect face region of the candidate with **skin colour detector.** The probability distribution over the colour of skin pixel in RGB space is modeled as a single Gaussian with full covariance. A corresponding Gaussian distribution with large variance is estimated for background pixel, and Bayes theorem is applied to obtain an image of the posterior probability that each pixel is skin. Skin blob detection is performed

to the skin probability image at each level. A face region is declared at local maxima in the DOG response with positive response above threshold, and corresponding high skin probability. The approximate scale of face is obtained. The pose based **face rendering** is performed by applying **3-D geometric model** with multiple **texture maps**. The single training image is back projected onto ellipsoid to given texture maps, can obtained new view of head move with different pose rendered by transforming the ellipsoid and projecting the texture maps in figure 1 back into the image. Multiple texture map can accurate rendering on many poses and differing appearance (facial expression). Pose is estimate and normalized with multiple appearances in-plane rotation with 6-D vector corresponding to rotation, scale and 2-D translation. Distance between poses is computed by the dot product between a front-facing vector normal to the ellipsoid. The representation of face image suitable for person classification performed using edge-based descriptor. Model is learned with probability distribution for the finding specific character in shot with variant in expression and pose.

2. 3-D ellipsoid approximation based constellation model

The [17] presented an approach to locate the individual character frames of person with large changes in scale without using the temporal information. The approach work in two stages first is to data collected automatically by frontal face detection and clustering it given ellipsoid model for the character and the corresponding texture map. For the variation in shape and other part of appearance invariant translation model used. Then apply constellation model over each aspect of local maxima and search using image pyramid for the verification with likelihood of model. While training stage, choosing patches around interesting points in the texture maps. PCA based model by Gaussian with diagonal covariance use for variation the data. The second stage is for verification generated by constellation model by assuming affine camera and four position of the corresponding ellipsoid to determine the pose. A gray scale image of ellipsoid model is rendered in the estimated pose for comparison against the input image.

3. Generic model with discriminative detector

In [18], work presented for automatic detection and identification of individual in unconstrained consumer video with a minimal number of labeled faces as training data.

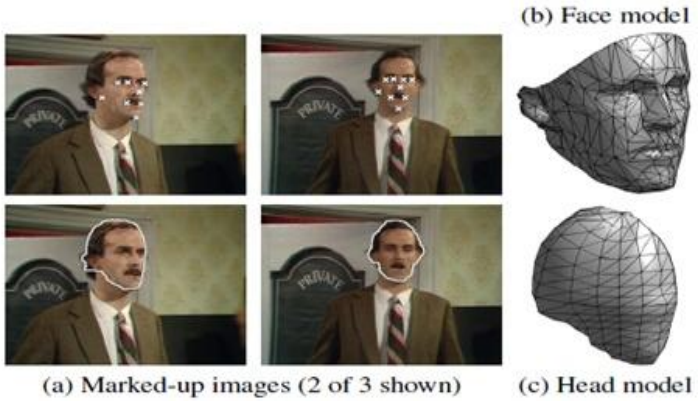


Fig 2. Modeling of face and head

First area has captured the 3-D appearance of the entire head, rather than just the face region and second area has to combined discriminative and 'generative' approaches for detection and recognition. The delivered approach consists of three parts-i) to built 3-D model of individual's face and head demonstrated in figure 2. The approximation of images of the head has to be rendered in novel view with few training images. ii) The pose has detected and estimated of individual using a trained tree-structured classifier. iii) Initial estimates of pose are refined, and identify verified using a generative approach and employing edge features and chamfer matching to give robustness to lighting and expression changes.

B. Part-(KLT TRACKER) Based

The facial image is divided into several regions and for each region, features are extracted using KLT tracker and compared with other region.

1. Face Sequence Matching in Large-scale Video Databases

The presented work in [19] uses face detector by J. Sivic *et al.* to locate human faces in every frame. Faces of the same person are associated together by tracking the covariance affine regions over time. Another approach is Kanade-Lucas-Tomasi(KLT) applied to every frame to track the interest points in a shot. A pair of face region indifferent frames is linked by comparing the number of tracked points that passed these face regions to the total points in both regions. Local Binary Pattern(LBP) features to represent the extracted faces,

LBP is invariant to monotonic changes in illumination and quickly computed. Each face is represented as a feature points, each face track described the distribution of the faces of one person in the feature space. The mean vector is used to represents the face tracks, it is nearly the approximation of the first principal component that corresponds to the direction of maximum variance of data. After the extraction and representation process are completed, the face tracks are organized into databases for the matching phase. From the given input face track, the similarity between input track and each track in the databases is estimated and rank list is returned according to the similarity score. The idea based on angle distance between two subspace, cosine distance is used to measure the similarity between two mean vectors.

2. Automatic Naming of character in TV video

[21] work in three threads, first processing of subtitles and script to obtain proposals for the characters in the video, by using fan web-site script of the video obtained in HTML format for human use. A "dynamic time warping" algorithm was used to align the script and subtitles in the presence of inconsistencies. The two texts were converted into a string of fixed-case, un-punctuated words. ; Second, processing of video to extract face tracks and accompanying descriptors, and to extract descriptors for clothing: The face detection with frontal face detector by Viola and Jones on every frame of video. The face is track for each shot with Kanade-Lucas-Tomasi Tracker .The output in the form of set of point tracks which are used to established relation between pair of faces within shot. Shot changes are automatically detected using colour histogram between consecutive frames. Output of face detector and tracker provides approximate location and scale of the face. Facial feature localized with nine facial features are located : left and right corner of each eye, the two nostrils and tip of the nose and left and right corner of mouth shown in figure 3 . To locate the feature generative model of the feature position combined with discriminative model of feature appearance using mixture of Gaussian tree. The affine transformation is used for geometrical normalization of pose variation. The two descriptor are used i) SIFT descriptor and ii) a simple pixel-wised descriptor. Euclidean distance is applied for computing distance between pair of face descriptor. With clothing appearance character is predicted in relative position using colour histogram. YCbCr colour space use for de- correlation. Speaker detected using significant lip motion. The third is combination of subtitle /script alignment and speaker detection given a number of face tracks identified with high probability i.e. min-min distance between descriptors.

(a) Face detections in original frames (b) Localized facial features

Fig 3.face detection and facial feature localization

3. Face Retrieval from Large Video Datasets

In [24] ,authors proposed the method for face retrieval in large video datasets. The faces of the same person appearing in individual shots are grouped into a single face track by using a reliable LBP tracking method. The retrieval is done by computing the similarity between face tracks in the databases and the input face track. The similarity between two face tracks is the similarity between their two representative faces. The representative face is the mean face of a subset selected from the original face track. The method tested on TRECVID large scale video database of 370 hours.

4. Face track finding in video

Thanh Duc Ngo *et al*.[25] presented a method for detecting face tracks in video in which each face represent an individual. The Kanade- Lucas-Tomasi(KLT) tracker to track interest points throughout video frames and each face track is formed by the faces detected in different frames that shear a large enough number of track points. For the face grouping: normally, the interest points (figure 4) are selected according to textured criterion that also found in facial regions. Track the points using motion model with the assumption that there is no much difference in features appearance and position. The match has decided by counting the number of shared points and total number of pair of faces in the frame. For the effect of false detection tracked points of only two consecutive frames considered. The intensity variance has been to measure the effect of illumination changes using flash light detector, which detect brightness of frame increases suddenly. For the effect of occlusion, a match between two faces in different frames which track number of track points that pass through the faces should be large enough compared to the total number of points inside the faces.

Fig 4. Interest point for tracking face

5. Real time person identification

[20] presented a person identification system which involves detection, tracking and recognition. Frontal face detector based on Viola-Jones is enhanced by Lienhart *et al***,** and is moderate fast. Face tracking with Relevance vector machine **(RVM)** tracker of William et al. that implemented a regularized kernel-based face tracker using radial basis function (RBF). The face recognition uses model of M. Everingham et al., which locate 7 facial features and four additional features are added at the center of the both eye, mouth and nose. The face region defined by facial features is normalized with canonical face to reduce the effect of scale and out of plane rotation of the head. Affine transformation is computed between canonical feature and facial features. To classify the faces a random- ferns classifier with 40 ferns of 17 levels and simple comparison performed between two elements of descriptor F, where F chosen random when building tree.

C. Color Based

1. Finding people in repeated shots of the same scene

In [26], authors proposed for finding all occurrence of a particular person in shot with changes in scale, pose and partially occlude. For the identification individual hair and clothing stays same throughout the sequence. The method having two stages, first for identification of individual by clustering frontal face detections using color clothing information. It has done with computing color histogram and group using single linkage hierarchical agglomerative clustering algorithm. Second a color based pictorial structure model for finding the occurrences of each person when frontal face detection was missed. The pictorial structure model is corresponding to three parts hair, face and torso region. The appearance of each modeled as a Gaussian mixture model (GMM) with k=5 components in RGB color space. For the common background have been computed part likelihoods. Prior shape probability has been calculated by 2D Gaussian distribution with mean and covariance. For minimizing matching cost, Maximum a posterior (MAP) has been used. The research will put on extension by improving the torso color models and considering multiple detection hypotheses with reference to author.

2. Face image retrieval in video sequence using lifting wavelets transform

Chon Fong Wong et al.[27] presented framework based on lifting wavelets transformed feature extraction. The Algorithm contained two stages; first stage is to extracts the key frame using color histogram method and second face detection and recognition. The face detection and recognition was further divided into four steps. The AdaBoost learning is use for face detection, which detects the face regions from key frames using For decease the illumination effect, facial region is normalization using cascade of classifier by performing histogram equalization which also reduce the computational time. Two level Lifting wavelets transformation for extracting facial features, lifting step means, any discrete wavelets transform (DWT) or two band subband filtering with finite sequence of simple filtering steps. Discriminate analysis LDA (Linear Discriminate Analysis) over PCA is applied to extract discriminate feature vector for the normalization. The last part means testing part i.e. the query image which is also a pre- processed through the face detection and extracting the features components, is compare with the feature vectors from the key-frames using the similarity measure i.e. Nearest Neighbor(NN) simple classifier which determine its existence in the video

3. Face Region Detection in MPEG Video

In [29], authors proposed three stages method for face region detection from compressed MPEG color video.

The input MPEG video minimal decoding and DCT of luminance and DCT PC or chrominance (Cb , Cr) (i. e. hue and saturation) the stage 1 – macroblock classification based on chrominance using human skin tone characterization which is generated by color station and Bayesian formula applying for minimum cost decision rule the classification an average chrominance which is minimum for false skin tone detection. In stage 2- detecting face region in macroblock mask image generated from stage1 which involve shape construction i. e. face detection frontal and nearly frontal faces for more accuracy in detection binary template matching is used on macroblock mask image. At last, Stage3 – verification of face detected bared on energy distributed of PCT coefficient correspond to vertical , horizontal and diagonal edge. The interceded form also district applied to MPEG-7 using DCTCS value of chrominance block and DCT coefficient of luminance block.

D. Multi clue (Fusion) Based

1. Character Identification in feature length films using global face –name matching

Proposed approach for the identification of character in feature length films using global face and name matching [22]. Face detected in video and clustered into groups of characters using Earth mover's Distance (EMP) for measure the distance of face tracks. Name affinity built from script. Speaking face tracks to build the face affinity network. Name and face association using matching vertices between two graphs. Multi-view face tracker to detect and track faces on each frame of the video, multi-view face exemplars in a track and finally cluster into groups corresponding to characters and built the face affinity network. Speaking face track detected using Region-of-interest and SIFT points are extracted and matched between current face image and the previous image. Face represented using Locally Linear Embedding (LLE) for dimensionality reduction. The distance measured between face track using spectral clustering and K dominant clusters from all the detected faces. The minimum distance makes them be treated as the same person due to the partial similarities. K-means clustering is performed to group the scattered face tracks which belong to the same character. The noise clear from clustering results with pruning method the marginal points which have low confidence belonging to the current cluster. The face-name association performed using affinity network in their own domain. Various application of the proposed work such as character relationship mining, character –centered browsing.

2. Speaker retrieval for TV show video

Proposed retrieval of speaker is in TV show programs using their names as the query [23]. The framework divided into two parts, visual processing part that focuses on shot segmentation and clustering of the same person with different occurrences together and locate their faces spatially and identify association part which focuses on solving the ambiguities between present faces and who and when speaks information. For the face detection, multi-view frontal face detector implemented in OpenCv by Viola and Jones used on every frame. The face is tracked by mean shift color tracker using color histogram or kernel density estimate of model and target image. The remaining part related to association of faces to the speaker because face tracker tracks all faces in a face image is not clearly specified.

3. Multimodal Person Search and Retrieval

Towards person Google: multimodal person search and retrieval [28] is a new search and retrieval approach. Content based on multimedia retrieval systems have been automatically indexing and retrieval multimodal person from video. The system work separately for audio and video segmentation, feature extraction and segment matching. The feature exaction from audio using mel frequency cepstral coefficient (MFCC). The temporal characteristics of audio data within segment reduce using multivariate Gaussian distribution. The audio segments are compared by applying Bayesian information Criterion (BIC) on computed the distance between two segments. For visual analysis face region determined based pn pupil position and anthropometric model. To handle the illumination changes, statistical normalization methods are applied globally and locally. The extracted features were reduces using PCA and the features were matched with database using Euclidean distance. The multimodal fusion consists of audio and visual score. Score has normalized with different characteristics at last the score fusion performed. It is fusion of face detection and speaker segmentation for person retrieval.

4. Face indexing system for actor –based video service in an IPTV environment

Jae Young choi et al. [30] proposed system for automatic system for indexing faces of actors. It consists of two parts, first to construction of FR engine using web and second is video face recognition. Using internet connect of STB, actor names are retrieved from online drama and movie information provider. For the second part, faces are clustered using colour histogram for computing each video frame for grouping the same subject into single subject cluster and face images of different subjects included in different clusters. The Hierarchical agglomerative clustering (HAC) has used for clustering, which terminates on threshold. The extraction of colour face features performed by the query face images (RGB) converted into different colour space like YIS or HSV images and extract features from these colour component vector creations. The extracted features from different colour components are concatenated. The face recognition has performed using weighted colour feature fusion, which dealing with several defective face images in a cluster which may contain variation in terms of viewpoint, illumination and compression artifacts. The elements penalty based Minowski, includes standard deviation from sample of feature vectors.

5. Estimate discriminant coordinates

In [31], authors proposed the face under pose variation of expression and illumination detection by Kernel PCA (kPCA) and discrimination for set and the rectification of canonical pose performed by running. Each SVM over image and performed affine transformation which best map detected points tp canonical

features. kPCA performed to reduce dimension and LDA to project data into space that for the discrimination task. The Nystram Approximation performs on kPCA dataset because its too large opt as kernel matrix. Finally the clustering is used to clear up the semi-super used database with error and new discrimination provide better representation of identification and re-cluster it in an modified K-means clustering.

6. Audio and visual information

Zhu Liu and Yao Wang [32] was investigated major cast detection in video by using both speaker and face information. The approach involves three steps. First step is speaker boundaries detection and clear speech extraction using GMM classifier. It has been covered the comparison of GMM classifier and SVM classifier. Simulated results are generated by GMM is better as compare to SVM. Then speaker segmentation and clustering using GMM distance metric with divergence. The second step to face detection in still image using fast template matching which detect multiple faces from still image. The face tracking and clustering involve. Video shot segmentation with distance of color histogram and face tracking within each shot using face template. Third and last step, integrated speaker and face correlation matrix i.e. speaker face –correlation matrix.

7. Local features and Statistical-Structural learning

This an integration of statistical and structural information that uses the local feature constructed from coefficient of quantized block transforms which is use in video compression. Under quantization and performing statistical histogram of the local features treated as vectors and similarity measure [33]. The image is decomposed into subarea called as local feature using quantized block transforms. Quantized coefficients of block transform are used for construction of local features and description called as feature vector. Ternary feature vector (TFV) structure from the collection of same order transform coefficient boring transformation blocks. Statistical information compared using TFV histogram based on 0^{th} and 4^{th} transform coefficient which represent different types of information about local feature. Structural description of pattern is represented by sub area histograms. Three aspects for pattern retrieval 1^{st} is the set of local features which is robust from perceptual point of view is not selected arbitrarily but by adjusting the quantization level of block transform. 2^{nd} size of selected feature histogram and last is scope of structural information.

8. Video Shot Retrieval for Face Sets

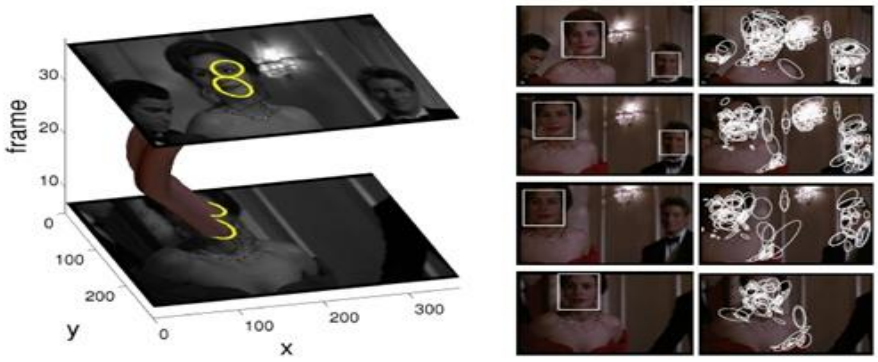


Fig 5. region track

The work suggested by J Sivic et al[14] for frontal face detection has been performed on every frame of the movie. This is achieved by first running a general purpose region tracker called affine covariant region tracker (fig. 5) by J. Sivic *et al* this tracking algorithm can develop tracks on deforming objects. Resultant, person's face can be tracked with significant pose variation and expression changes, but tracking is done offline. A single shot contains hundreds of frames with possibly one or more detected faces and the detected faces generally connected by several region tracks. A single-link agglomerative grouping strategy is used to merges face detection into larger group starting from most closest(most connected).

Each face sets (face-track) are described by a collection of five affine transformed local spatial orientation fields based around facial features. The entire set represented as a single distribution over local feature descriptors. The facial features(left and right eyes, tip of nose and center of the mouth) localized allows to local face descriptors and affinely deform support regions to normalize for pose variation. A probabilistic part-based "constellation" model of faces is used to modal the joint position (shape) and appearance of the facial features. Facial features are located by searching for joint position of the features which maximizes the posterior probability of the feature positions and appearance. Each face in the set is represented as a collection of five local overlapping parts(SIFT descriptors)placed at the detected feature locations(eyes, mouth, nose and midpoint between eyes)[14].

9. Film Character Retrieval In Feature-Length Films

Ognjen Arandjelovie *et al.* [16] has proposed approach to recognize all the frontal faces of a character in the movies or situation comedy shown in figure 6, given a small number of query faces. The recognition method based on cascade of processing steps that normalized the effect of changing imaging environment, particularly in three areas suppress the background of face, pose refinement to optimize the registration and used robust distance to a sub-space to allow for partial occlusion and expression changes.

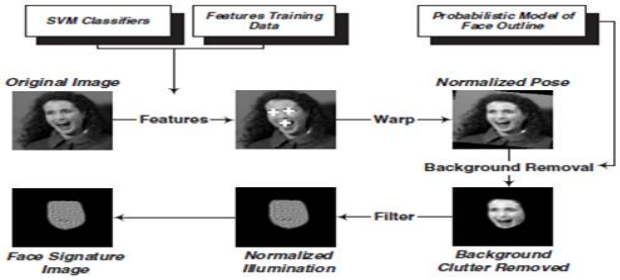


Fig 6. Face representation

Ognjen Arandjelovie *et al.* [16] considered content-based multimedia retrieval setup to retrieval and rank by confidence from film shots. A query consists of user choosing the person of interest in one or more key-frame. A face detection stage has performed by local implementation of 3D object detection based on correctly detection of both eyes and mouth, are visible. The proposed approach consists of computing the low distance in numeric value, a distance, expressing the degree of belief that two images belongs to same person and computing a series of transformation of the original image.

IV. ANALYSIS

The table 2 provides the techno-critical analysis of the human face retrieval approaches. The process involves certain inter related task to achieve the objective. The following table shows the technical details of concern approaches to accomplish the work on different media files. The Table 3 shows the details examination of the available approaches with respect to input- output parameter and results performed.

The detail analysis (Table 3) presents the involvement of different input parameters, advantages, limitations of human face retrieval approaches from video. The some approaches provides encouraging results, but under the certain assumptions. The all the presented approaches has the different objectives, in accordance with application.

V. PERFORMANCE EVALUATION AND RESULTS

The analysis of the different approaches concerned in tables, the approaches have different objectives to deal in real life problems. The major concern of the analysis is to evaluate the performance. For the evaluation as a part, consider the objective that the human face image retrieve from video database.

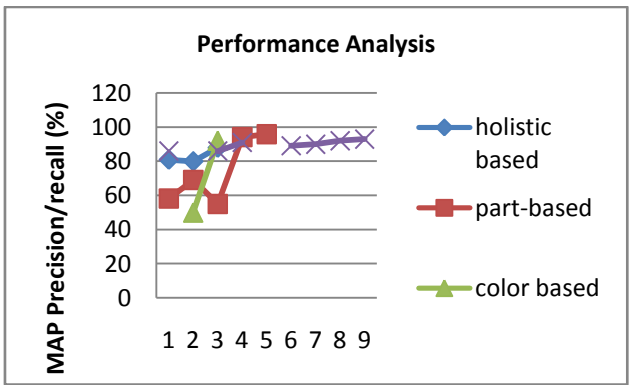
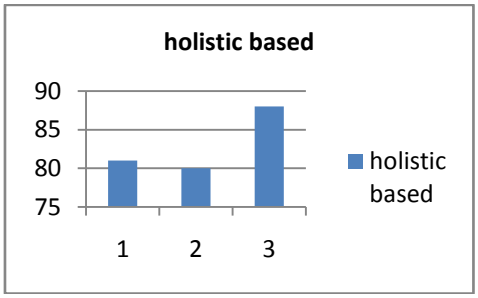


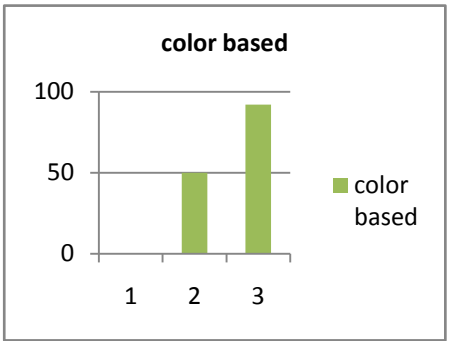
Fig 7 MAP performance analysis (precision/ recall)

We consider the MAP (mean average precision/recall) measure to estimate performance of available approaches. The Fig 7 shows the average performance analysis (precision/recall) for the different categorization of human face retrieval on various available approaches. The data has been considered from the results and evolution of approaches, which are mentioned in table 3 as status of results. The available data varies with

respect to database, size, specification, etc., however the performance measured on one scale. It is the possibility of changing the status of results with respect to changed in data or process techniques.
The graph drown in fig. 7 includes all the four categorized to human face retrieval with all mentioned techniques/approaches. Multi clue has large number of approaches proceed because of wide scope of implantation. Total 9 approaches consider for the concerned. Whereas, holistic and color based got modest scope i.e. three for each, with respect to the current age. The part-based approached has considered 5 approaches under the category with average and stable precision rate.



(a)



(b)

Fig 8. Mean precision/recall rate of holistic (a) and color based (b)

Fig 8. Shown the mean precision/recall rate, for the evolution of holistic and color based approaches. The holistic contained 3 approaches with average 83.33% of precision/ recall rate (table 3). The highest mean precision is 88% and lowest is 80%. Whereas, color-based includes highest precision rate with 92% and average 71%, because the first method of this category has not provides precision/recall rate.

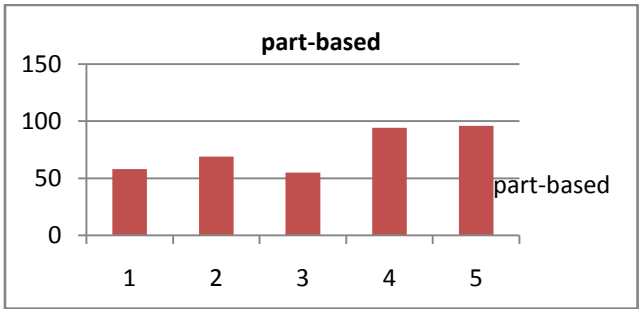


Fig 9. Mean precision/recall rate of part-based

The part-based has the highest mean precision rate of 95.78% with lowest 54.97% under the control or uncontrolled database and specification. It has average precision/recall rate 74.41% of 5 different approaches under various conditions. The fig 9 provides the representation of the same graphically. The Multi-clue approach has the highest number of approaches because of wide varieties of combination. In future, it can be identified sub categorized according to (table 1) or by considering other widespread and significant specification. Fig 10 described the 7 different performances out of 9 available approaches for mean precision rate. The two approaches did not have precision rate, so the average precision rate in accordance with the 7

available approaches. The highest precision rate is 93% and lowest is 85.99% and the average precision rate 89.59%.

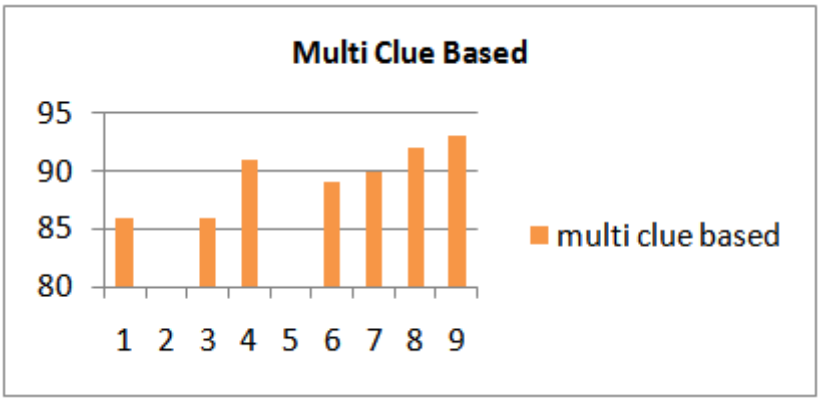


Fig 10. Mean precision/recall rate of Multi-Clue based

Viola and Jones detector performed better results as compared to other for detection in video and images. Adaboost performed fast and more accurate results. The KLT tracker performed good results but it varies to illumination changes. The track points have advantages over KLT and performed better in various conditions. The histogram matching and probability distribution finds good results for recognition, but it varies due to unfavorable situation occurred.

The skin tone color for detection and recognition may cause computational complexities. The results of retrieval better due to face track of grouping of similar faces. The fusion of information is not provides matured results. The less work has been done the large-scale video database with optimum speed and accuracy.

VI. CONCLUSION

The proposed paper provides an optimized review and analysis of different available approaches for human face image retrieval. With the help of above study, the decision has been made for making the framework which provides feasible solution. The work has been extended for the various objectives such as searching, browsing and indexing. The useful investigation has sum up with better results in terms of accuracy and speed under the assumption that human face retrieval from video databases.

Table2. Techno-critical Analysis and exploration of different face retrieval approaches

Approach Category	Approach of implementation	Steps used in Approach									
		Face detection	**Tracking**	**Feature extraction**	**Normalization**	**Matching**	**Representation**	**Classification**	**Training data**	**Recognition**	**Other techniques**
Holistic (3-D model based)	1. 3-D model with texture maps[15]	Skin color detector	Deformable region tracker	colour of skin pixel in RGB space	3-D geometric Model	Threshold	3-D model with multiple texture maps	Edges	3-D geometric model	probability distribution	face rendering
	2. 3-D ellipsoid approximation based constellation model[17]	Affine invariant clustering	Did not used tracker	Principal Component Analysis	3-D Constellation model	Threshold, ranking by likelihood	3-D ellipsoid face model	constellation model	Did not used temporal information	ellipsoid model is rendered for comparison	face rendering
	3. Generic model with discriminative detector[18]	Discriminative detector using 3-D rendered model	Did not used tracker	Adaboost	3-D model of individual's face and head	Chamfer distance	3-D model	Edge detection	minimal no. of labeled faces as training data	Chamfer distance and confidence measure	trained tree-structured classifier
Part-based(KLT tracker)	1. KLT with LBP[19]	Viola and Jones detector	Kanade-Lucas-Tomasi(KLT)	Local Binary Patter(LBP)	Eigenvector, mean-zero	Cosine distance	Mean vector	Local Binary Pattern(LBP)	Did not used training data	Ranking	370 hrs of TRECVID news video
	2.Combined visual-textual model[21]	Viola and Jones detector	Kanade-Lucas-Tomasi(KLT)	Nine facial features on position	SIFT descriptor and simple pixel-wised descriptor	Euclidean distance	Computing descriptor of local appearance	Quasi-likelihood	descriptors for clothing	High probability i.e. min-min distance	descriptors for clothing
	3. face-track matching[24]	Viola and Jones detector	Face track: track points	LBP (local binary patterns)	No grey level normalization needed	k-Faces	Face track: LBP (local binary patterns) feature	Clustering based distance	Face track extraction	Euclidean distance between two mean faces	TRECVID dataset
	4. Track point based method[25]	Viola and Jones detector	Kanade-Lucas-Tomasi(KLT)	Interest points	normalization not needed	count the no. of shared points and total no. of pair of faces	features appearance and position	total number of points inside the faces	Did not used training data	Compared track interest point	**Face grouping**(Face track)
	5.IDU system[20]	Viola and Jones detector enhancement by Lienhart	Kernel-based tracker using radial basis function(RBF)	13 facial features	Affine transformation	Set of posteriors (Max-max or max-sum)	13 facial features	Random-ferns classifier	Cascade face detector	Pictorial structure model(9)+4 additional facial features	40 ferns of 17 levels
Color based	1. Color based pictorial structure model[26]	FloatBoost learning or Viola and Jones detector	single linkage hierarchical agglomerative clustering algorithm	Gaussian mixture model (GMM) with k=5	2D Gaussian distribution with mean and covariance	Score matching with Max. posterior probability(MAP)	Modeling with GMM	Maximum a posterior (MAP)	sequence of photograph taken over a short period of time	Pictorial structure	**Face grouping using color histogram**
	2. Lifting Wavelet Transform Feature transform[27]	Viola and Jones detector	Color histogram	Lifting Wavelet Transform (LWTFE)	PCA	Nearest Neighbor classifier	LDA	AdaBoost	Did not used training data	Cosine distance	Nearest Neighbor(NN) simple classifier
	3. Compressed-domain approach[29]	Human skin tone	Skin tone color chrominance	macroblock mask image	macroblock mask image	energy distributed of DCT coefficient of vertical, horizontal and diagonal edge	macroblock mask image	macroblock classification based on chrominance using human skin	binary template matching	energy distributed of DCT coefficient	DCTCS value of chrominance and DCT coefficient of luminance block
Multi clue based	1. Affinity network based global face-name matching[22]	Earth mover's Distance(EMP)	Multi-view face tracker	spectral clustering and K dominant clusters	K-means clustering	Matching face tracks with (Euclidean distance)	Locally Linear Embedding(LLE)	K-means clustering between face track	Did not used training data	Gaussian Mixture Models(GMM)	Faces grouped in each shot
	2. Fusion of audio-visual information[23]	OpenCv based Viola and Jones detector	Mean shift color tracker	kernel density estimate	color histogram	Histogram matching	Eigenface or local facial features(SIFT)	Did not used	Did not used training data	Histogram matching	shot segmentation and clustering
	3. Multimodal fusion system[28]	Face region determined by pupil positions and anthropometric face model	Did not used tracker	PCA	z-score	Euclidean distance	PCA	pupil position and anthropometric model	multivariate Gaussian distribution	Score level fusion(product, sum, min, max)	score fusion performed
	4. Audio and visual information[32]	Face template matching	Skin color distribution with average face template	Skin tone color	Generating Average face template	Average face template matching(maximum Matching Value)	Face template	Face track clustering	Average face template	χ^2 Distance in color histogram	Shot segmentation by color histogram
	5. Estimate discriminant coordinates[31]	Nystrom Approximation	Did not used tracker	kPCA and LDA	Cholesky Decomposition	Low likelihood with threshold	Cluster	Modified K-means clustering	Did not used training data	Nearest Neighbor	Merging clusters
	6 Spatial–temporal system[14]	Propagation probability	affine covariant region tracker	Principal Component Analysis	local face descriptors and affinely deform support regions	Histograms compared using χ^2	overlapping five SIFT descriptors	Histogram	Principal Component Analysis	probability density function	Face exemplars (Region tracking)
	7. local features and Statistical-Structural learning[33]	Searching best subarea	Did not used tracker	discrete cosine transform (DCT) block	lower-order AC coefficients	Histogram matching	Histogram of Ternary feature vector (TFV)	TFV Histogram using single subarea	Training images	Histogram matching	Statistical-Structural learning Approach
	8. Cascade approach[16]	SVM based detector	Did not used tracker	SVM	Background removal, band-pass filtering, rank matching	Threshold	Face registration using facial features	Background removal	SVM	Comparing face signature image	illumination is normalized by using band-pass filtering
	9. Weighted feature fusion scheme[30]	Face clustering : Hierarchical agglomerative clustering(HAC)	faces are clustered using colour histogram	Color component (YIQ) face features	z-score	Nearest Neighbor classifier	Color component vector creation	Hierarchical agglomerative clustering (HAC)	Did not used training data	Weighted color features fusion	Shot segmentation by color histogram

The techno-critical analysis (Table 2) presents the involvement of different major and minor aspects of human face retrieval from video. The detection involved the location of face image if present. Whereas, tracking approach, tracks the face image (template) or points in video sequences. The different features of face images were extraction and represented after normalization. The face image has classified from the background or other effects. However the matching and recognition are the final parameter for the identification. Since sometime training images need to made an final face image representation. According to the user requirements, the different parameter involves to make a complete face retrieval process successful.

Table 3. Detail investigation of different approaches for face retrieval

Approach Category	Approach of implementation	Input Parameter	Intermediate input	Advantages	Limitation	Objectives and constraints	Status of Results	Output	Approach of matching
Holistic(3-D model based)	1. 3-D model with texture maps[15]	minimal labeled faces as training data	3-D model	•Wide variation of pose and appearance considered •Efficient search	• Extra execution time for 3-D model • Effective for only color video	Person identification in video with wide variation in pose	Pose covers in video around +/- 60°	Identities location of particular persons	Face image to video
	2. 3-D ellipsoid approximation based constellation model[17]	minimal labeled faces as training data	3-D model	•Wide variation of pose and appearance considered	•Variation in pose and facial expression •Unsupervised model	Find every instance of a particular person in video	True positive up to 80% and false positive below 10 %	find every instance of a particular person in video	Face image to video
	3. Generic model with discriminative detector[18]	minimal labeled faces as training data	capture 3D appearance	•Edge-based measure for pose refinement and recognition •Robustness to lighting, deformations by expressions	• Effective for only color video • Error correction technique not define	Identification of person in video with minimal number of training data	recall level of 50% and precision around 80% for all characters	automatic detection and identification of individuals in video	Face image to video
Part(KLT tracker) based	1. KLT with LBP[19]	Face image query	Face-tracks	•Large-scale database(TRECVID) •Minimum computational cost	• Error correction technique not define	Speed and accuracy for large video database	mean –cos method obtained the MAP at 58.13%	matching face sequences	Face image to video
	2.Combined visual-textual model[21]	Script, face track, clothing color	Track points	•combination of textual and visual information	•mechanism for error correction •accuracy decreases in combined approach	High precision, character naming in TV video	accuracy obtained is around 69%	labeling appearances of characters	Face image to video
	3. face-track matching[24]	Video	Face tracks	•Large-scale database(TRECVID) •Minimum computational cost	•Method did not work efficiently in real life video	Speed and accuracy for large video database	MAP(precision/ recall) is 54.97%, computation cost is good	Face retrieval from large video datasets	Face image to video
	4. Track point based method[25]	Video	Face tracks	•Recomputed and complement points in case of loss or degraded •Flash light detector to detect illumination changes	•Interest point did not generated for key frames •Re-computation required additional computation cost	Face tracks grouping , accuracy	total detected tracks are 94.17% (i.e. 711/755)	detecting face tracks in video	Face image to video
	5.IDU system[20]	Face image query	11 facial features	•Track up to 5 people in real time •Better classification performance than min-min classifier	• Auditory feedback via headphones • Automatically add new person	Real time person identification ,speed and accuracy	95-98% precision at 20% recall	person identification	Face frame to video
Color based	1. Color based pictorial structure model[26]	Sequence of photograph taken over a short period of time.	Clothing color	•Discover people by clustering frontal face detector •Re-detect people not facing camera using clothing and hair	•Spatial layout of clothing colors •torso color model implement •Multiple face detection	All occurrence of particular person in a sequence of photograph taken over short period	53 missed detections (out of 59) are correctly filled-in	finding people in repeated shots	Face image to image data
	2. Lifting Wavelet Transform Feature transform[27]	for face image	key-frames	•small memory cost, and robust face detection and recognition performance	•face alignment is not included •expected associated computation cost increases	Speed and accuracy for video sequences	average faces image retrieval accuracy of 50 top ranks	face image retrieval in video sequences	Face frame to video
	3. Compressed-domain approach[29]	inverse quantized (DCT) coefficients of MPEG	binary template	•detect face regions direct from MPEG video streams •provides better results than real time system	•Other information in MPEG video stream not included •More accuracy in detection and recognition	Human face region detection in MPEG	92% detection rate in different sizes, frontal and side-view faces	locations of the detected face regions	Face image to video
Multi clue based	1. Affinity network based global face-name matching[22]	video and film script	names and clustered face tracks	•global face-name matching using affinity network •build face-name association	•Gender and context information for refine results •Achieve better accuracy and completeness by organizing on shot level	Global matching, accuracy	Accuracy about 86% with better precision/recall rates	identifying characters in feature-length films	Face frame to video
	2. Fusion of audio-visual information[23]	names as the query	frame sequence	•Increase the retrieval performance in TV show video with audio speaker recognition. • It does not required training steps for specific person	•Details framework did not provided •Accuracy not calculated	Speaker retrieval for TV shows video	Quantitative Results not specified	retrieving speakers in TV	Face frame to video

Multi clue based	3. Multimodal fusion system[28]	Query by example	Audio and video	• multimodal person search and retrieval using voice and face characteristics	• different query paradigms and relevance feedback techniques incorporate into system • detection step (face detection and speaker segmentation) analyze the performance	Indexing and retrieval of audiovisual data	mean recall values are R={49;58;66}% and precision values are p={24;29;33}%	content based multimedia retrieval system	Face frame to video
	4. Audio and visual information[32]	speaker information and face information in tracks	Face track	• Temporal and spatial appearance • Face tracking with template	• Association of talking face detection • Name tag detection	Major cast detection in video, accuracy, association	Results are simulated individual for speaker and face information is 91% Appox.	automatic generate list of major casts in a video sequences	Face frame to video
	5. Estimate discriminant coordinates[31]	Face image in news, captions		• Fixing incorrect elements within cluster	• Accuracy not calculated • Reject threshold	Names and faces in news	Quantitative Results not specified	Names with face in news	Face image to video
	6.Spatial–temporal system[14]	Face image query	Face-tracks	• Distinctive inter-personal and intra-personal matching • Representation of face sets reduces matching problem	• Intra-shot matching to clustering over entire movie • Exclusion of negative exemplars at run time	Video shot retrieval using face sets	The 98.1% precision in intra-shot matching	shots containing particular people	Face image to video
	7. local features and Statistical-Structural learning[33]	Face image	Histogram TFV	• Local features of face subarea • statistical-structural learning with histogram	• Accuracy not calculated • Extra execution time • Complicated approach	Retrieval time, training time and storage space	Rank-1 CMS score (%) above 90%	Face retrieval	Face image to video
	8. Cascade approach[16]	small number of query faces	rank by confidence	Removes the distortions-pose, background, illumination and partial occlusion	• Flexible model for learning • Incorporating temporal information •	• Recognized all frontal faces	High recall rates over 92% achieved	all the frontal faces of a character in video	Face image to video
	9. Weighted feature fusion scheme[30]	high number of training images	clustered face images	• video face recognition are unsupervised • low computational complexity	• On-line learning for dynamic adjustment	Face indexing in IPTV environment , accuracy	precision of more than 90% and recall of more than 93%	face indexing system	Face frame to video

REFERENCE

[1]. Ming-Hsuan Yang, David J. Kriegman and Narendra Ahuja, *Detecting Faces in Images: A Survey*, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO. 1,pp-36-58, JANUARY 2002.

[2]. Xinzhu Wang ,Yantao Tian, Shuaishi Liu, Jinsong Li, Cheng Peng ,*Face Detection and Tracking Algorithm in Video Images with Complex Background*, Proceedings of the 2010 IEEE International Conference on Robotics and Biometrics ,pp.- 1206-1211, December 14-18, 2010.

[3]. Ukil Yang, Minsung Kang, Kar-Ann Toh and Kwanghoon Sohn, An *Illumination invariant skin-color model for face detection*, Proceedings of the IEEE International Conference, 2010.

[4]. Lutz Goldmann, Ullrich J. Mönich and Thomas Sikora, *Components and Their Topology for Robust Face Detection in the Presence of Partial Occlusions*, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL.2, NO.3, pp-559-569, SEPTEMBER 2007.

[5]. WU Lushen, WU Peimin, MENG Fanwen, *A Fast Face Detection for Video Sequences*, IEEE Second International Conference on Intelligent Human-Machine Systems and Cybernetics, pp-117-120, 2010.

[6]. Raphael Sznitman and Bruno Jedynak, Active *Testing for Face Detection and Localization*, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 32, NO. 10, pp- 1914-1920, OCTOBER 2010.

[7]. Li Sun and Guizhong Liu, "Visual Object Tracking Based on Combination of Local Description and Global Representation", IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 21, NO. 4, PP. 408-420, APRIL 2011.

[8]. Jae Young Choi, Yong Man Ro and Konstantinos N. Plataniotis, "Boosting Color Feature Selection for Color Face Recognition", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 20, NO. 5, PP. 1425-1434, MAY 2011.

[9]. Zak Ston, Todd Zickler and Trevor Darrell, "Toward Large-Scale Face Recognition Using Social Network Context", Proceedings of the IEEE | Vol. 98, No. 8, PP.1408-1415 August 2010.

[10]. K. Ruba Soundar and K. Murugesan, "Preserving global and local information – a combined approach for recognising face images", IET Computer Vision, Vol. 4, Iss. 3, pp. 173–182, 2010

[11]. Stefano Berretti, Alberto Del Bimbo, and Pietro Pala, "3D Face Recognition Using Isogeodesic Stripes", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 32, NO. 12, PP. 2162-2177, DECEMBER 2010.

[12]. Jae Young Choi, Wesley De Neve, Konstantinos N. Plataniotis and Yong Man Ro, "Collaborative Face Recognition for Improved Face Annotation in Personal Photo Collections Shared on Online Social Networks" , IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 13, NO. 1,PP.14-28, FEBRUARY 2011.

[13]. Ran He, Wei-Shi Zhengand Bao-Gang Hu, "Maximum Correntropy Criterion for Robust Face Recognition", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 33, NO. 8,PP. 1561-1576, AUGUST 2011.
[14]. J. Sivic, M. Everingham, and A. Zisserman, "Person Spotting: Video Shot Retrieval for Face Sets", Proceeding of ACM International conference on image and video retrieval, (CIVR'05), pp.226-236, 2005.
[15]. Mark Everingham and Andrew Zisserman, "Automated person identification in video" , Proceedings of ACM International Conference on Image and Video Retrieval (CIVR'04), page 289--298, 2004.
[16]. Ognjen Arandjelovie *et al* , "Automatic face recognition for film character retrieval in feature-length films" , In proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp-860-867, Jun 2005, Print ISBN: 0-7695-2372-2
[17]. Mark Everingham and Andrew Zisserman , "Automated visual identification of character in situation comedies" , ACM 17th International Conference on Pattern Recognition (ICPR'04), Vol. 04, pp 983-986, 2004, ISBN:0-7695-2128-2.
[18]. Mark Everingham and Andrew Zisserman, "Identifying individuals in video by combining generative and discriminative head models", Proceedings of the tenth IEEE International Conference on Computer Vision (ICCV), Vol. 2, pp- 1103 – 1110, Dec 2005, ISSN: 1550-5499, Print ISBN: 0-7695-2334-X.
[19]. Hung Thanh Vu, Thanh Duc Ngo, Thao Ngoc Nguyen, Duy-Dinh Le, Shin'ichi Satoh, Bac Le Hoai, Duc Anh Duong, "Fast Face Sequence Matching in Large-scale Video Databases", IEEE International Conference on Image Processing (ICIP 2011), Sep 2011.
[20]. Nicholas Apostoloff and Andrew Zisserman, "Who are you? – Real time person identification" ,Proceedings of the British Machine Vision Conference(BMVC) , 2007.
[21]. Mark Everingham et al. , "Hello! My name is… Buffy" –Automatic Naming of character in TV video" , Proceedings of the British Machine Vision Conference(BMVC) , 2006
[22]. Yi-Fan Zhang, Changsheng Xu *et al*, "Character Identification in feature length films using global face – name matching" IEEE TRANSACTION ON MULTIMEDIA, Vol. 11, No. 7, pp-1276-1288, Nov 2009.
[23]. Yina Han *et al.* , "Speaker retrieval for TV show video by associating audio speaker recognition result to visual faces", Proceedings of 2nd International conference the K-Space , 2008.
[24]. Thao Ngoc Nguyen, Thanh Duc Ngo *et al*., "An Efficient Method for Face Retrieval from Large Video Datasets" proceeding of ACM International Conference on image and video retrieval (CIVR), 2010, ISBN: 978-1-4503-0117-6.
[25]. Thanh Duc Ngo *et al.*, "Robust face track finding in video using tracked points", IEEE International Conference on Signal Image Technology and Internet Based Systems, SITIS '08, pp-59-64, Dec 2008, Print ISBN: 978-0-7695-3493-0
[26]. Josef Sivic *et al*., "Finding people in repeated shots of the same scene", Proceedings of 16th British Machine Vision Conference(BMVC), Vol. 3, pp-909-918, 2006.
[27]. Chon Fong Wong *et al*., "Face image retrieval in video sequence using lifting wavelets transform feature extraction", IEEE Proceedings of the Ninth International Symposium Consumer Electronics (ISCE 2005)., pp-167-171,Jun 2005, Print ISBN: 0-7803-8920-4.
[28]. Lutz Goldmann , Amjad Samour and Thomas Sikora "Towards Person Google: Multimodal Person Search and Retrieval" , In Proceedings of International conference SAMT, pp.292-295 , 2007.
[29]. Hualu Wang and Shih-Fu Chang, "A Highly Efficient System for Automatic Face Region Detection in MPEG Video", IEEE TRANSACTION ON CIRCUIT AND SYSTEMS FOR VIDEO TECHNOLOGY, Vol-7, No. 4, pp-615-628, August 1997.
[30]. Jae Young Choi, Wesley De Neve and Young Man Ro, "Towards an automatic face indexing system for actor –based video service in an IPTV environment", IEEE TRANSACTION ON CONSUMER ELECTRONICS , Vol. 56, Issue 1, pp-147-155, Feb 2010.
[31]. Tamara L. Berg et al. , "Names and faces in the news" , Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04) - Volume 2, pp-848-854, July 2004, ISSN: 1063-6919 , Print ISBN: 0-7695-2158-4.
[32]. Zhu Liu and Yao Wang, "Major cast detection in video using both speaker and face information" , IEEE TRANSACTION ON MULTIMEDIA, Vol. 9, No. 1, pp-89-101, Jan 2007.
[33]. Daidi Zhong and Irek Defee , " Face retrieval based on robust local features and statistical structure learning approach" , EURASIP Journal on Advances in Signal Processing , Volume 12,2008.

Faizur Rashid. "Critical Analysis of Various Approaches for Automatic Human Face Image Retrieval." IOSR Journal of Engineering (IOSRJEN), vol. 09, no. 08, 2019, pp. 26-39.