

Anchor Person Detection using Haar-Like Feature Extraction from News Videos

Brindha M.
Research Scholar
Department of Computer Science
PSGR Krishnammal College for Women

R. Amsaveni
Assistant Professor
Department of IT
PSGR Krishnammal College for Women

ABSTRACT

The human face and facial feature extraction play a key role in person identification in the areas of video surveillance and access control on security reason. In this research work news video is taken for anchor person detection. Detecting anchor person from news videos give the time distribution to news readers and provide editorial support to journalist to find videos related to the particular person. First the video is converted into frames and the shot change detection algorithm is used to find scene changes and store the image in the database. Two different algorithms are used to find scene change detection such as color based shot detection and edge based shot detection. From these result, edge based shot detection performs well in finding shot changes more accurately. Second, segment the still image into skin region and non-skin region by using skin-color model based on its size and shape face region is identified. Third step, facial features like location of eye, nose and mouth are extracted to recognize face variations through haar-like features. It provides a possible ways to locate the positions of eyeballs, mouth centers, midpoints of nostrils and near and far corners of mouth from face image. This approach helps to extract features on human face automatically and improve the accuracy of face detection. Finally, anchor person is detected from news video by extracting SURF features of the given image. Experimental results show methods used in this research could locate facial features from face exactly and quickly.

General Terms

Detection, boundary, feature extraction, etc.,

Keywords

Anchor person, SURF, haar-like feature, edge-based, color histogram.

1. INTRODUCTION

In recent years, automatic detection and recognition of human faces from still image and videos are an active research area. Even though people are good at face identification, recognition of human face automatically by a computer is a very difficult task. Face recognition has been widely applied in credit-card verification, teleconference, and criminal identification systems and so on. Face recognition is subjective by many complications, such as the differences of facial expression, the variety of postures, and the light directions of imaging, angle and size [1], [2]. Even for the same people, images taken in different surroundings may be distinct. The problem is so difficult that the success in the field of automatic face recognition by computer is not as gratified as the finger prints visualization. Facial feature extraction has become the main issue in automatic recognition of human faces from still images and videos. Detecting the

basic feature like eyes, nose and mouth accurately is necessary for most face recognition methods for accurate identification.

A complete face detection system [3] could identify facial features and localize in faces by facial expression and pi-point features from a given image. Most of the facial expression and facial recognition system work on the assumption that the location of the face presented within the specific region of a frame is known [4]. This kind of identification is only suitable for scenes with the same background; conversely, complex background faces are localized before any of the recognition is performed. This issue is analyzed in this work, proposed face detection system will work based on color and shape information also then locate eyes, mouth and nose through haar-like features.

The remaining paper is formatted as follows; section II gives the news videos structure and categories. Section III describes the proposed work for the anchor person detection framework. Section IV demonstrations the experiments and its results in detail. Finally, section V presents the conclusion and future work.

2. NEWS VIDEOS

News is packaged information about current events happening somewhere else. News moves through media like, postal systems, based on word of mouth, broadcasting, printing, and electronic communication [5]. The common topics for news reports contains politics, business, and war, as well as athletic events, strange or unusual events, news about celebrities, government announcements, about royal ceremonies, public health, laws, criminals and taxes have been dubbed news since ancient times.

Humans show evidence of a nearly worldwide desire to learn and share news from to a different place, which satisfy by talking to each other, sharing and traveling. Technical and social developments are often driven by government communication and surveillance networks, have increased the speed with which news can spread, as well as subjective its content [6], [7].

2.1 News Categories

2.1.1 Business

Each clip provides a report on one business story made available as stories break. It includes a selection of business videos for each week, such a business look ahead, looking at what to expect in the markets that week, and more evergreen features looking at money and tips for customers and people with interesting careers.

2.1.2 National

In the National news category, each news segment focuses on one story, either spot news, breaking news, features or sidebar

stories. The average of eight national news clips is created and made available on weekdays, and up to three on weekends.

2.1.3 World news

The largest breadth and depth of breaking news around the world, including entertainment, consumer, technology, health, business, lifestyle, quirky and sports news which is approximately of 25-30 clips per weekday and fewer on weekends were broadcasted.

2.1.4 Sports

Each one-to-two-minute clip gives an extended report on sports story, formed and made available as stories break.

2.1.5 News Explainers

Event-based stories chronicle the news of a city. Such as this bill was passed, this person was hired, that person was fired and etc. News Explainers make sense of the news rather than just telling you what happened. News Explainers divide the news like why or how it happened.

2.1.6 Major Breaking News

Cities are drenched with everyday news stories such as traffic jams, fires and etc. But Major Breaking News has a much bigger impact on a city or a region based on the news. Enormous storm is an easy example of this because they tend to make life complex for whole regions.

Most of the news videos have similar and well-defined structures [8]. The Chaisorn et al. [9] have studied the structure of news videos and noticed:

“The news video naturally begins with several highlight video sequences that give a brief introduction of the upcoming news to be reported. The main body of news includes a sequence of stories organized in terms of different geographical interest, such as international, national, regional and local, and in broad categories of social business, political, entertainment and sports. Every news story usually begins with an anchorperson and most broadcasts include reports on Sports, Finance or Weather.”

Figure 1 exemplifies the structure of a typical news video. Also the ordering of news items may vary slightly from the broadcast station to station, but all have a similar structure and news categories [10], [11].

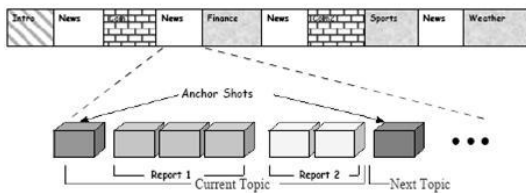


Fig 1 General news structure

3. ANCHOR PERSON DETECTION FRAME WORK

The proposed framework includes three different phases; following diagram (Fig 2) demonstrates the proposed work of anchor person detection. In the first phase, the given video is taken as input and edge-based shot detection method is used to extract scene change frames and stored in the database. Second phase, human face and upper body are detected based on haar-like features. Third phase, shape modeling is used as a key factor in case of anchor person detection. Head and shoulder method is used to find the anchor person detection.

SURF features are extracted by identifying the anchor person image more accurately.

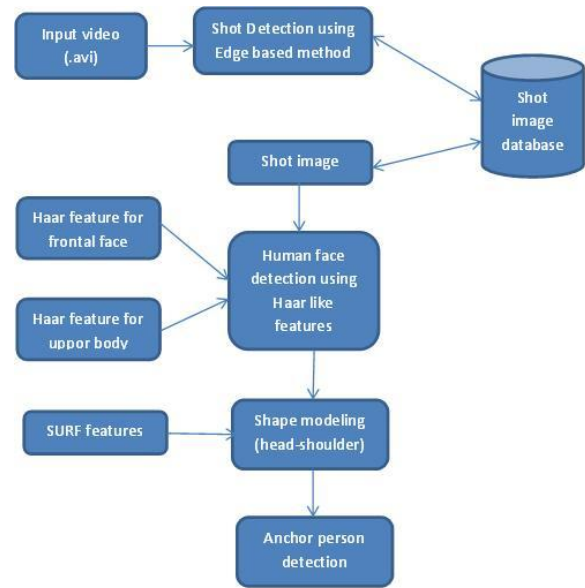


Fig 2 Frame work for anchor person detection

3.1 Shot detection

Color histogram features have been used broadly related to shot detection field. The easiest way of applying the color histogram technique is to calculate the mean value of the RGB for the frames and group similar frames to detect a shot from video sequences. In this work color histogram algorithm is used to find scene change detection based on the RGB mean value. Most approaches to video segmentation make use of the color information in the video. The limitations of color histogram in video segmentation are the problems of illumination variation and motion-induced false alarm. So the edge-based shot detection is implemented to solve the problem of illumination and motion estimation. The result of the shot detection is given in the following section.

3.2 Haar Like Features

Haar-like features are digital imaging features which are working only on image intensities (i.e) RGB values of each and every pixel in an image. A Haar-like feature considers neighboring rectangular regions at a definite location in a detection window, each region pixel intensities are summed up also calculates the difference between these sums. Then it is used to categorize subsections of an image. In this work image database with human faces are used. For all faces the region of the eyes is darker than the region of the cheeks is a common observation. So a common haar feature for face detection is a set of two adjacent rectangles that lie above the eye and the cheek regions. The region of these rectangles is defined relative to a detection window that acts like a bounding box to the target object of the human face.

3.3 Surf Features

Speeded Up Robust Features (SURF) is a robust local feature detector used in this work to extract relevant features and descriptors from an image. It uses an integer approximation to the element of the Hessian blob detector, which can be computed very quickly with an integral image. The Haar wavelet response around the point of interest is summed and used as features; again, these can be calculated with the assistance of the integral image. This approach is preferred

over its predecessor due to its pithy descriptor length (64 floating point values). In SURF, a descriptor vector of length 64 is created using a histogram of gradient orientations in the local neighborhood around each key point in an image. The function returns extracted feature vectors, also called as descriptors, and their equivalent locations, from a binary or intensity image. It originates the descriptors from pixels surrounding an interest point in an image. Pixels represent the match feature specified by a single-point location also specify the center location of a neighborhood and the method used for descriptor extraction depends on the class of the input points.

3.4 Shape Modeling

Blob detection is like a segmentation problem. Conversely, blob detection is based on chromatic cues, but it depends on the camera's quality and color contrast between background and foreground [12], [13]. For the real time tracking purposes, it is also likely that errors gathered during the course. In order to alleviate the blob tracking process, the system should be adjusted from time to time basis. In this work, shape cues are used to meet the requirements [14]. Based on object silhouette peoples can sometimes identify the head and shoulder region only. Though the foreground is always in motion so the shape is keeps on changing, but it is not in every case so the shape is analyzed successfully for segmentation purpose. The following figure (Fig3) has two foreground silhouettes to be analyzed [15]. In the left image, the shape feature is well-built and the head region and shoulder region can be easily detached based on the foreground shape information, while in the right image, the shape feature is not so observable.

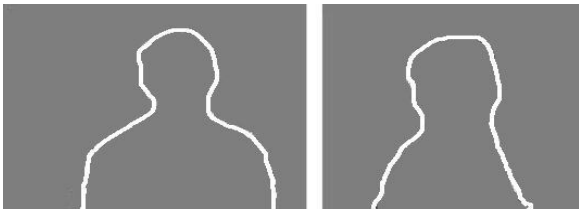


Fig 3 Shape model

To solve this problem, a shape recognition module is added into the shape analysis procedure. To define those shapes with strong shape features as recognized head- and -shoulder shape. The shapes outside this category are named non-recognized shape. If a foreground shape fits into the recognized canonical category, sure that can locate the head region and shoulder region with a high reliability only on shape information, then head blob and shoulder blob are located with shape cues.

Though a computable definition of recognized or non-recognized category is hard to give and may depend on specific algorithm [16] of shape analysis, comprehensiveness is not the requirement of this work. Moderately, because the purpose of shape analysis is to alleviate the region-based blob tracking; define the recognized shape category as a limited size. That is, do not mean that all the shapes are included in the segment the foreground into the head and shoulder regions with certain simple shape analysis algorithm. Instead, only those with strong shape features are included in the category to make sure that if a shape is accepted as a recognized one, then the segmentation output based on the shape analysis is highly consistent.

If a foreground shape is a recognized shape, by definition its shape features is used to segment the foreground into head region and shoulder regions, or in other words, to segment the foreground map $f(x, y)$ into a support map $s(x, y)$ of head

blob and shoulder blob. A containing rectangle based algorithm is designed to segment a foreground map $f(x, y)$ into two blobs by setting a vertical threshold t :

$$s_1(x, y) = \begin{cases} f(x, y) & \text{if } y > t \\ 0 & \text{otherwise,} \end{cases}$$

And

$$s_2(x, y) = \begin{cases} f(x, y) & \text{if } y < t \\ 0 & \text{otherwise,} \end{cases}$$

If the operator $S(s_k(t))$ and $S(rect_k(t))$ to represent the size of support map k and size of its containing rectangle at the thresholding t respectively, i.e.,

$$S(s_k) = \sum_{(x,y) \in I} s_k(x, y)$$

$$S(rect_k) = (y_t - y_b)(x_t - b)$$

Where I represents the image pixel set, y_b, y_t, x_b, x_t then the final segmentation thresholding T is chosen as

$$T = arg \min_t \left\{ \sum_{k=1}^2 [S(rect_k(t)) - S(s_k(t))] \right\}$$

This algorithm is illustrated in Figure 4. Note the coordinate system and the horizontal thresholding line that segments the foreground into the head and shoulder regions. The purpose of this algorithm is to approximate the foreground map with two containing rectangles and choose the segmentation that minimizes the approximation error.

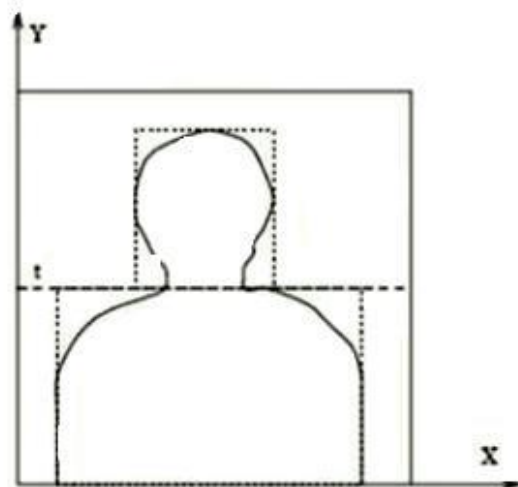


Fig 4 Blob region

4. EXPERIMENTS AND RESULTS

4.1 Data Collection

The news videos are taken from YouTube website (www.youtube.com). Four different videos are downloaded from two different channels like NDTV and TV9 news with two different time duration such as 10 minutes and 5minutes respectively.

4.2 Edge Based Shot Detection

Edge-based shot detection approach is used to find scene changes from given news video. Two frames are compared based on the segmented edge information, if the average distance between same segmentation frames is taken as input

and stored in database as detected shot image. The following figure (Fig5) illustrates the edge-based shot detection result.

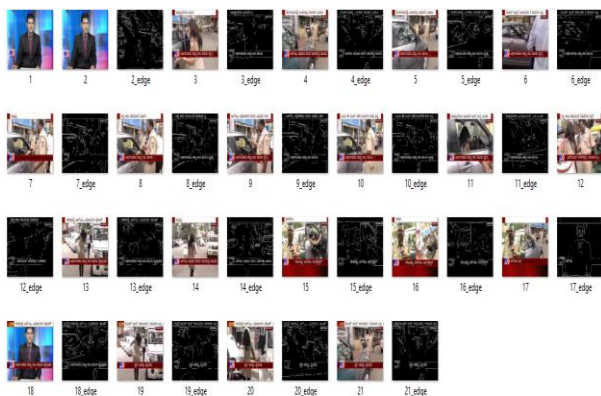


Fig 5 Edge based shot detection.

4.3 Haar Cascade Based Frontal Face And Upper Body Detection

First, a boosted classifier working with haar-like features is trained with a few sample views of a particular object's face, called positive samples, which are scaled to the same size of 20x20, and negative examples - arbitrary images of the same size are trained.

After a classifier is trained, it can be applied to a region of interest in an input image. The output is "1" if the region is likely to show the face otherwise the result is "0". To search for the object in the whole image move the search window across the image and check every location using the classifier. It is designed so that it can be easily "resized" in order to find the objects of interest in various sizes, which is more effective than resizing the image itself. So, to find an object of a new size in the image the scan process should be done several times at different scales. Similarly, the upper body detection is done for a given image. Following figure (Fig6) shows the result of frontal face and upper body detection using haar-cascade based approach.

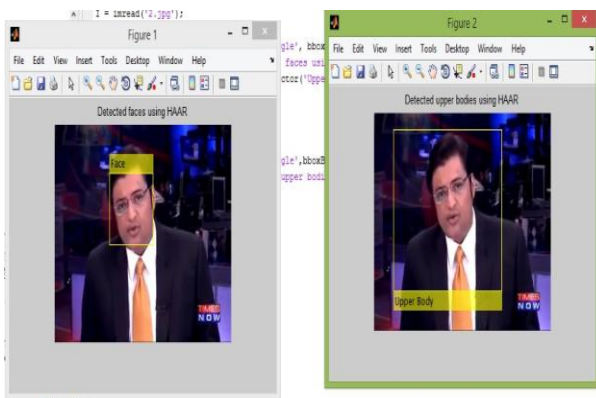


Fig 6 Frontal face and upper body detection using Haar-cascade based approach

4.4 Facial Feature Extraction And Location

From the detected face facial features like big eyes, small eyes, right eye, left eye, nose and mouth are detected and confined. The following figure (Fig7) gives the illustration of facial feature and location.

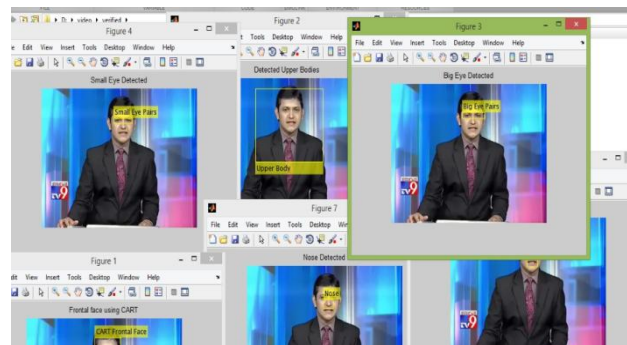


Fig 7 Pointing of facial features

4.5 Anchor Person Detection Using Shape Modeling

The anchor person was detected based on blob and SURF features. First the face portion of the image is taken and resized then SURF features are extracted from face image if the feature points are reached the thresholding level then the image contain anchor person in it. The following figure (Fig 8) demonstrates the detected anchor person shot and the figure (Fig 9) shows the non-anchor person detected shot.



Fig 8 Anchor person shot

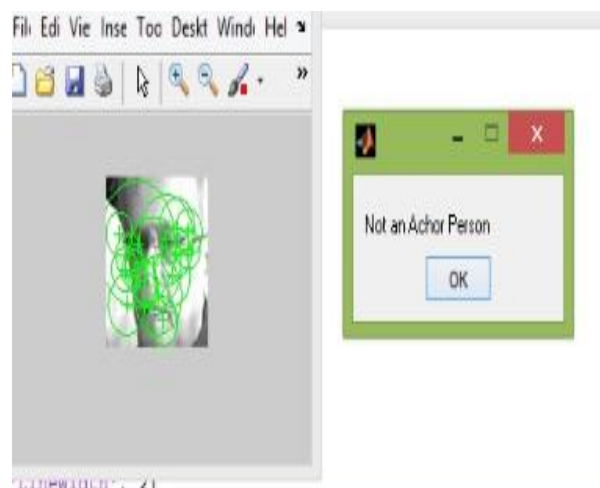


Fig 9 Non Anchor person shot

5. CONCLUSION

This research work present anchor person detection from news videos. It includes two different phases; first phase news video is converted as scenes, shots and frames. There are two different algorithms are used to find scene change detection like color histogram based shot detection and edge based shot detection. Based on the detected shots anchor

frame is identified by using shape modeling techniques. Cascade based Haar-like features are used to find an anchor person using Haar like features and SURF features. The experimental result demonstrates that the proposed work performs well in finding the anchor person from news videos accurately and quickly. With regard to future work, it's not limited up to shape analyze using cascade based Haar-like feature. Enhance to find all visual features and texture features from a news video to help journalist and news readers for editorial purpose.

6. REFERENCES

- [1] R. Castagno, T. Ebrahimi, and M. Kunt, "Video segmentation based on multiple features for interactive multimedia applications", *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):562-571, 1998.
- [2] N. Oliver, A. Pentland, and F. Berard, "LAFTER: A real-time lips and face tracker with facial expression recognition", In *Proceedings of International Conference on Computer Vision and Pattern Recognition*, S.Juan, Puerto Rico, June 1997.
- [3] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions and wavelet packet analysis", *IEEE Transactions on Multimedia*, 1(3), 1999. Tavel, P. 2007 *Modeling and Simulation Design*. AK Peters Ltd.
- [4] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions and wavelet packet analysis", *IEEE Transactions on Multimedia*, 1(3), 1999. Forman, G. 2003. *An extensive empirical study of feature selection metrics for text classification*. *J. Mach. Learn. Res.* 3 (Mar. 2003), 1289-1305.
- [5] T. S. Chua, S. F. Chang, L. Chaisorn, and W. Hsu, "Story boundary detection in large broadcast news video archives -Techniques, experience and trends," in *Proceedings of the 12th ACM International Conference on Multimedia*, pp. 656–659, October 2004.
- [6] P. Joly, J. Benois-Pineau, E. Kijak, and G. Qu'enot, "The ARGOS campaign: evaluation of video analysis and indexing tools", *Signal Processing*, vol. 22, no. 7-8, pp. 705–717, 2007.
- [7] A. E. Abduraman, S. A. Berrani, and B. M'erialdo, "TV program structuring techniques: a review", in *TV Content Analysis: Techniques and Applications*, 2011.
- [8] J. M. Gauch, S. Gauch, S. Bouix, and X. Zhu, "Real time video scene detection and classification", *Information Processing and Management*, vol. 35, no. 3, pp. 381–400, 1999.
- [9] L. Chaisorn, T. S. Chua, and C. H. Lee, "A multi-modal approach to story segmentation for news video", *World Wide Web*, vol. 6, no. 2, pp. 187–208, 2003.
- [10] C. Ma, B. Byun, I. Kim, and C. H. Lee, "A detection-based approach to broadcast news video story segmentation", in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 1957–1960, April 2009.
- [11] L. Chaisorn and T. S. Chua, "Story boundary detection in news video using global rule induction technique", in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '06)*, pp. 2101–2104, July 2006.
- [12] H. Misra, F. Hopfgartner, A. Goyal et al., "Tv news story segmentation based on semantic coherence and content similarity", in *Proceedings of the 16th international conference on Advances in Multimedia Modeling*, pp. 347–357, 2010.
- [13] A. Goyal, P. Punitha, F. Hopfgartner, and J. M. Jose, "Split and merge based story segmentation in news videos", in *Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, pp. 766–770, 2009.
- [14] L. Shen, L. Bai, "A review on Gabor wavelets for face recognition", *Pattern. Anal. & Applications*, vol. 9, issue 2, pp 273-292, Sep. 2006.
- [15] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition", *IEEE Trans. Image Process.*, vol. 11, No. 4, pp. 467–476, April. 2002.
- [16] X. Huang, S. Z. Li, and Y. Wang, "Shape localization based on statistical method using extended local binary pattern", in *Proc. International Conference on Image and Graphics*, pp.184-187, 2004.