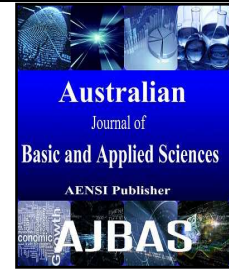




AUSTRALIAN JOURNAL OF BASIC AND APPLIED SCIENCES

ISSN:1991-8178 EISSN: 2309-8414
Journal home page: www.ajbasweb.com



Scene Change Detection Approaches Over UT Interaction Dataset

Mithun Chakaravarti, Mithul Mathivanan, G.S. Mahalakshmi, J. Betina Antony and S. Sendhilkumar

Dept of Computer Science and Engineering Anna University, Chennai 600025

Address For Correspondence:

Mithun Chakaravarti, Dept of Computer Science and Engineering Anna University, Chennai 600025.
E-mail: mithun2595@gmail.com

ARTICLE INFO

Article history:

Received 04 December 2015

Accepted 22 January 2016

Available online 14 February 2016

Keywords:

Frame extraction, UT interaction, gradual scene change, scene change detection

ABSTRACT

Scene change detection is the initial step in video processing applications. Proper scene change detection helps many applications like video segmentation, action recognition and video matching to proceed error-free. Though many methodologies are presented in the literature, many algorithms were scene change type dependant. This paper presents a survey of existing scene change detection algorithms. These algorithms were applied over the human UT interaction video datasets. The results show the requirement for gradual scene change detection of human activities.

INTRODUCTION

The tremendous development in surveillance cameras’ usage, storage capacity and internet conferencing facilities has laid platform for usage of videos in all fields including intelligence surveillance. Recording of human activities during meetings and in public places for future investigation has been increased. Automated action recognition software needs to identify humans from background followed by recognition of actions between humans. When the action done by human becomes continuous, there is a need for initial segmentation of individual sequence followed by detection of change. There are three different ways of identifying the action performed by a human in videos.

- Identifying a single person’s action.
- Identifying action between two humans using predefined examples (Similar to supervised classification).
- Classifying action between two humans without prior knowledge.

Various challenges exists while detection the change in videos. It might be due to complex scenarios, noise and occlusions. In order to overcome these difficulties, appropriate care should be taken at segmentation level according to the chosen video dataset. Videos involving complex scenarios can have dense background and many unnecessary objects blocking the required human portions. When the input dataset contains noise, the effectiveness of the algorithm to be applied will become less. Hence prior elimination of noise from the videos is must.

When videos contain other challenges like shadows, multiple persons, two actions performed simultaneously by two different persons, the most important step is to identify the real and imaginary humans (shadows). Much research has been done already for action recognition and scene change detection in videos with single person. But videos involving multiple persons require more effort. The scenario becomes more difficult when humans perform actions in a continuous manner.

Scene change detection is carried out as the key step in most of the video processing applications. The change in scene usually has two variants – abrupt change and gradual change. Abrupt change is sudden transfer of scenes in the video. During abrupt change all the objects in the scene will be switched with other objects and the

Open Access Journal

Published BY AENSI Publication

© 2016 AENSI Publisher All rights reserved

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

To Cite This Article: Mithun Chakaravarti, Mithul Mathivanan, G.S. Mahalakshmi, J. Betina Antony and S. Sendhilkumar. Scene Change Detection Approaches Over UT Interaction Dataset. *Aust. J. Basic & Appl. Sci.*, 10(2): 60-66, 2016

resultant will have different objects or background or person. During gradual change, the scene transforms slowly from one stage to another. The change transforms in a bit by bit manner. Major gradual changes include transitions like fade in, fade out and dissolve. When videos contain human interactions, the change in scene will mostly have gradual changes.

A pixel based approach was proposed by Xiaoquan Yi *et al.* to detect abrupt scene changes. It was carried out in two steps. The first step uses tested the mean absolute frame differences (MAFD) followed by normalized histogram equalization process in the second step. It clearly detects the sharp changes and camera motion (Yi, 2005). Other than pixel based scene change detection methods, many methods were proposed. Some detecting algorithms use likelihood ratio (Ford, M., 1997; Dugad, 1998), some use the mutual information between frames [5] and some others use the frequency domain correlation values (Porter, 2000; Vlachos, Theodore, 2000).

A histogram based approach was given by Nisreen *et al.*, Histogram correlation was used to detect gradual scene changes. Using a reference frame, the frames extracted from the video is compared. Based on the correlation value, detection of gradual scene change is carried out (Radwan, I. 2012; Meng, Jianhao, 1995).

Scene changes should also be detected from a compressed video. Jianhao *et al.*, presented an abrupt scene change detection algorithm which detects changes from a MPEG/MPEG-2 compressed video. Since the video is collection of compressed bit stream, minimal decoding becomes sufficient. This reduces the computational time of the algorithm. The additional advantage of this algorithm is that it can be applied in scene browsing and video indexing.

An algorithm called twi-difference algorithm (Liu, 2003) was developed to detect abrupt scene changes from videos. Inter-frame difference algorithm is taken as the base for this algorithm. It also compares frames based on pixels. It is sequence independent and reduces the computational cost. This twi-difference algorithm has no missing with close-to-zero or little false detection for the abrupt shot change. The video movement's influence on abrupt shot change detection is greatly eliminated in twi-difference algorithm. However more expansion is needed if gradual scene change needs to be detected.

Lassoued *et al.*, developed a new methodology for detecting action changes in videos using moments correlation (Lassoued, 2012). Silhouettes poses of humans in given videos were initially segmented and they are described using 2D Krawtchouk moments. These silhouettes were used to derive the cross correlation matrices and Kullback- Leibler distance. This is used to detect the action changes done by humans. This method is based on cross correlation measures between silhouette poses and global video descriptor.

Various other methods were also used with graphs. Scene segmentation is transformed into a graph partitioning problem (Rasheed, 2005). With each shot as a node, and their similarity determining the edges between the nodes, a shot similarity graph was constructed. Then using the motion (scene change) information and the shot similarity graph, scenes were segmented. The transition of a video from one scene to another scene was depicted using a transition graph (Yeung, 1998). In this graph, the connected sub graph corresponds to the scenes in the videos. A completely different method was given by Zhai *et al.*, (2006). They use Markov chain Monte Carlo to determine scene boundaries.

Scene change detection can also be used for many applications. Chung-Lin has employed scene change detection for segmenting the videos. It detects both abrupt and gradual scene changes. A static scene test and an intensity statistics model were developed to detect scene changes from videos. It works best for videos involving gradual scene changes (Huang, 2001). Scene change detection is also used for video compression. Ankita Chauhan has employed scene change detection algorithm over an uncompressed video and based on the scene change detected frames, video is compressed. A block based motion estimation (BME) is combined with scene change detection method to convert uncompressed video into a compressed video. This hybrid approach has greatly reduced the computational complexity without a drop in PSNR value (Chauhan, P., 2013).

In our work, various approaches which help in identifying the change in scenes are explained. It is experimented to check which algorithm or approach is best suited for scene change detection in videos involving humans' interaction with gradual changes. These approaches are tested over UT interaction videos and the results are observed.

MATERIALS AND METHODS

Research for scene change detection from videos is increasing in the recent days. It is expanding based on the requirement that the change should be detected based on the type of change involved in videos. Initially the video undergoes a process to extract the frames from it. Frames will be in image format and the collection of entire frames will result in original video. An algorithm is applied over the extracted frames in order to check the amount of change it has gone through from one frame to another frame. Based on the amount of change it has gone through, the algorithm determines the exact change position.

Here four different general scene change detection techniques were taken and employed over human UT interaction videos. They can be categorized as pixel based methods and histogram based method. Pixel based methods are

1. Image differencing method.
2. Image rationing method.

3. Fuzzy XOR based method.

A. *Pixel based Methods:*

An image is a function of two variables, say $f(i, j)$. Here i and j are integer values usually taking values from 0 to width/height of the image. For a binary image, $f(i, j)$ is a single value, either 0 or 1. For a gray scale image, its value is from 0 to 255. If the image is a color image, then it will be represented in *rgb* format. It stores the intensity value a pixel has. Two successive frames were taken and the amount of change it has undergone is checked based on the change in number of pixels. If there is no motion in video, then there will not be any change in the intensity values. Hence change in number of pixels will be nil. Various pixel based methods were taken and employed over the human UT interaction videos and its efficiency over them is checked.

1) *Image differencing method:*

It is the simplest pixel based method to detect change in between frames. Two successive frames indicate two images taken at successive times of same area. These two successive frames were compared using the difference in their intensity values. If there is no change in scene, then the difference value will be zero (0). Based on the amount of change the scene has gone through, the difference value will vary. For an abrupt change, the overall difference value between successive frames will be very high. Whereas for a gradual scene change, the resultant difference value will be moderate. The algorithm illustrating the image differencing method is given below.

Image Differencing method - Algorithm

Input: Frames from given video

Output: plot indicating the difference values.

```
begin
  for each frame
    read the frame and successive frame
    calculate the pixel values of both frames
    calculate difference between the pixel values
  end for
  plot the difference value
end
```

2) *Image Rationing Method:*

Similar to image differencing, image rationing also compares frames based on pixel values. Two successive frames were taken and their ratio is found out. If the images are same, then the ratio will take the value as 1. The ratio value will indicate the amount of change it has gone through between frames. When the ratio image is taken as the intermediate result, it will indicate the positions in the image that has undergone change. A continuous change in scenes of successive frames indicates motion in video. Thus continuous change in ratio value will determine the motion of the video. The image rationing method is described as pseudo code.

Image Rationing Method - Algorithm

Input: Frames from given video

Output: Plot indicating the ratio between successive frames.

```
begin
  for each frame
    read the frame and successive frame
    calculate the pixel values of both frames
    calculate ratio between the pixel values, i.e.,
       $I_r(x, y) = I_1(x, y) / I_2(x, y)$ 
  end for
  plot the ratio values
end
```

3) *Fuzzy XOR Method:*

It also a pixel based scene change detection method. Binary XOR operation is taken as the basic function and fuzzy XOR operation is used for scene change detection from videos. Considering black and white frames obtained from video, the pixels will take values either white or black. An XOR operation over these frames will yield an image which shows the pixels that are changed. Let p_1 and p_2 be the pixels from two successive frames, frame 1 and frame 2 at position (x, y) . The existence of change is indicated in the resultant image if either " p_1 is white and p_2 is black" or " p_1 is black and p_2 is white".

Similarly for gray scale and color video, fuzzy XOR operation is employed to get the scene changed pixels. Starting from the first frame, two successive frames were compared using fuzzy XOR operator. When there is motion in the video, there will be a hike in fuzzy XOR value between those two successive frames. Based on the increase in fuzzy XOR value, the type of scene change can be determined.

Fuzzy XOR based method - Algorithm

Input: Frames from given video

Output: plot indicating the fuzzy XOR values of frames

begin

 for each frame

 read the frame and successive frame

 calculate fuzzy XOR value using the XOR function between frames

 end for

 plot the obtained fuzzy XOR values for different frames

end

B. :Histogram Based Method

Histogram based methods are alternate for pixel based algorithms. Pixel based methods compare a specific pixel between two successive frames. Whereas histogram based methods are pixel independent. Each frame will have its own color histograms thus similar frames resulting in similar color histograms. From the given video, frames are extracted and histograms are generated for each frame. Then Euclidean distance is calculated between successive frames.

$$d(p, q) = \sqrt{\sum_{i=1}^n \sum_{j=1}^m (p(i, j) - q(i, j))^2}$$

For two successive frames p and q the Euclidean distance d(p,q) is expanded as,

Histogram based method – Algorithm

Input: Frames from given video

Output: plot indicating the comparison of histograms

begin

 for each frame

 read the frame and successive frame

 convert to gray scale image if it is rgb

 calculate the Euclidean distance between the frames

 end for

 plot the Euclidean distance based compared histogram values

end

RESULTS AND DISCUSSION

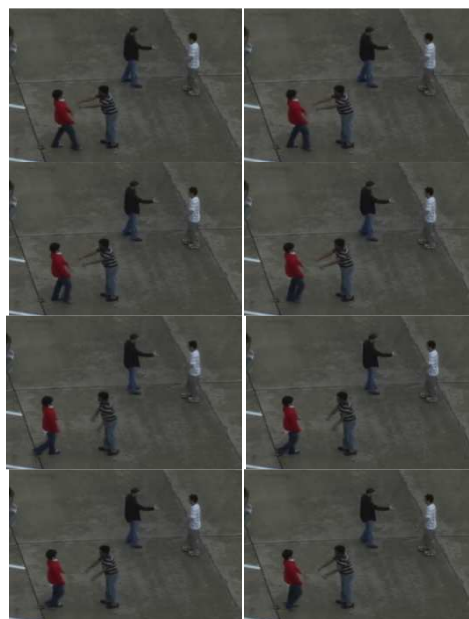


Fig. 1: Sample frames extracted from UT interaction video (clockwise from top-left).

The dataset used is the UT interaction dataset (Chaquet, M., 2013). It consists of six classes of human-human interactions: hand shaking, punching, hugging, pushing, pointing and kicking. It consists of more than two humans and in different clothing's. Initially frames were extracted from videos. A sample of extracted frames is shown in figure 1. It starts from top left corner and goes in clockwise direction, ending at bottom left corner. The difference in scene can be seen between the first frame and the last frame in figure 1.

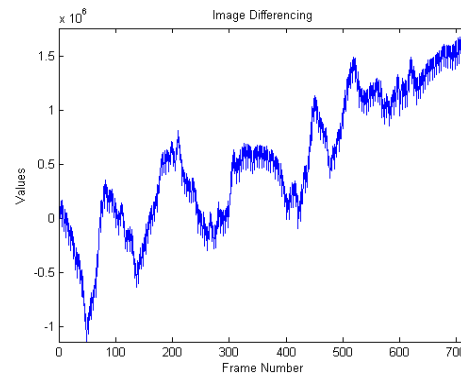


Fig. 2: Histogram showing image differencing values of an UT interaction video.

Figure 2 shows the histogram which depicts the image differencing values between two successive frames. Given video contains 720 frames and it is taken in x-axis. The values are given in y-axis. If the difference value is very high, there is sudden rise or sudden fall in the graph. It indicates the motion in the video. The difference value can also take negative values based on particular frames. Similarly figure 3 depicts the graph between the frames of a video and the ratio values between the frames. When there is no motion in video, the ratio value between two successive frames will be one. If there is scene change the image ratio value will be very high. The increase in image ratio value corresponds to the scene change position in the video.

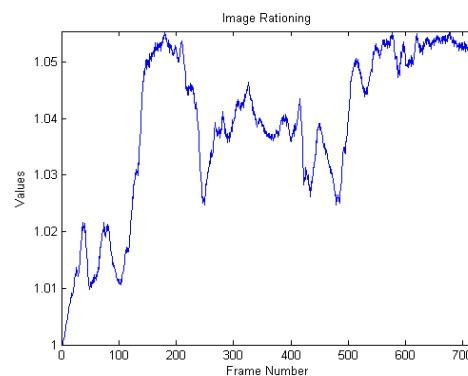


Fig. 3: Histogram showing image rationing values of an UT interaction video.

Fuzzy XOR operation is also applied to the 720 frames and its result is plotted as a graph as shown in figure 4. The frame number against the fuzzy XOR values is taken and the value describes the scene change for a non-zero fuzzy XOR value. If there is no change in the video, then the resultant will be zero i.e. a black pixel. Higher value indicates a complete scene change in video. Histogram based method's result is depicted in figure 5. As similar frames have similar histograms, the plot also has very close histogram values except few places. Perfect cuts can be exactly indicated using this histogram based method for scene change detection.

These pixel based and histogram based methods are compared using measures like recall, V, Precision, P and F-measure. For the given UT interaction video, 720 frames were initially extracted and the desired number of scene change in the video is 10. Correct, C corresponds to the exactly identified scene changes, M corresponds to the number of scene changes that were not identified. Some improper detection was also done and it is called as false positives, F. Of all the four methods, pixel based methods have almost found all the scene changes. Histogram based method have not found even an average number of changes. The reason is the scene location does not change thus making all the frames look alike. The false positives occur due to the slight change in light intensities between frames. Table 1 shows the evaluation carried out for both pixels based and histogram based methods for UT interaction dataset.

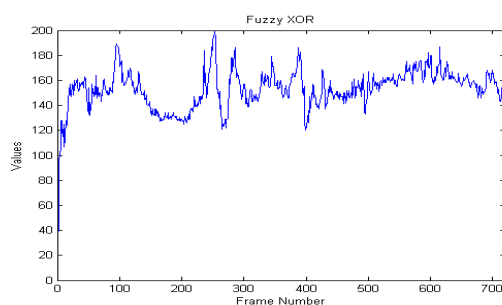


Fig. 4: Plot showing fuzzy XOR values for frames of an UT interaction video.

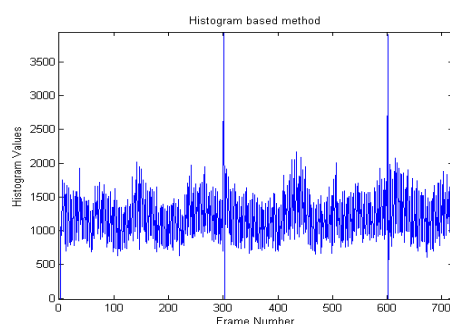


Fig. 5: Plot showing difference values after applying histogram based method over an UT interaction video.

Conclusion:

This paper has presented the efficiency of the scene change detection algorithms when applied over UT interaction video dataset. Initially frames were extracted from the given video (UT interaction video) and then four algorithms were applied over them. It can be seen from the obtained results that the existence of false positives has decreased the accuracy of the methodologies. This is because the changes in the background were also considered as scene change. It can be avoided by an initial step of segmenting the humans from the entire frame. Human specific features can be used for scene change detection in UT interaction dataset.

Table I: Comparison of various methods based on F-measure.

Method	No. of frames	Desired	Correct (C)	Missed (M)	False Positive (F)	Recall (V)	Precision (P)	F - Measure
Image differencing	720	10	8	2	4	0.8	0.667	0.727
Image Rationing	720	10	6	4	5	0.6	0.545	0.571
Fuzzy XOR based	720	10	9	1	3	0.9	0.75	0.818
Histogram based	720	10	4	6	6	0.4	0.4	0.4

REFERENCES

Butz, Torsten, and J. Thiran, 2001. "Shot boundary detection with mutual information." In Image Processing, Proceedings. International Conference on, 3: 422-425. IEEE.

Chaquet, M., Jose, Enrique J. Carmona, Antonio Fernández-Caballero, 2013. "A survey of video datasets for human action and activity recognition." Computer Vision and Image Understanding, 117(6): 633-659.

Chauhan, P., Ankita, Rohit R. Parmar, Shankar K. Parmar, Shahida G. Chauhan, 2013. "Hybrid approach for video compression based on scene change detection." In Signal Processing, Computing and Control (ISPCC), IEEE International Conference on, 1-5. IEEE.

Dugad, Rakesh, Krishna Ratakonda, and Narendra Ahuja, 1998. "Robust video shot change detection." In Multimedia Signal Processing, 1998 IEEE Second Workshop on, 376-381. IEEE.

Ford, M., Ralph, Craig Robson, Daniel Temple, and Michael Gerlach, 1997. "Metrics for scene change detection in digital video sequences." In Multimedia Computing and Systems, International Conference on, 610-610. IEEE Computer Society

Huang, Chung-Lin, Bing-Yao Liao, 2001. "A robust scene-change detection method for video segmentation." *Circuits and Systems for Video Technology*, IEEE Transactions on, 11(12): 1281-1288.

Lassoued, Imen, Ezzeddine Zagrouba, and Youssef Chahir, 2012. "A new method using Moments correlation for action change detection in video." In *IEEE International conference on Innovative Computing Technology*, 1(3): 174-178.

Lee, Seong-Whan, Young-Min Kim, and Sung Woo Choi., 2000. "Fast scene change detection using direct feature extraction from MPEG compressed videos." *Multimedia*, IEEE Transactions on, 2(4): 240-254.

Liu, Qianlei and Luxi Yang, 2003. "Twi-difference algorithm for video's abrupt shot change detection." In *Neural Networks and Signal Processing*, Proceedings of the International Conference on, 2: 1177-1180. IEEE.

Meng, Jianhao, Yujen Juan, and Shih-Fu Chang, 1995. "Scene change detection in an MPEG-compressed video sequence." In *IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology*, 14-25. International Society for Optics and Photonics.

Porter, Sarah, V. Majid Mirmehdi, and Barry T. Thomas, 2000. "Video cut detection using frequency domain correlation." In *Pattern Recognition*, 2000. Proceedings. 15th International Conference on, 3: 409-412. IEEE.

Radwan, I. Nisreen, Nancy M. Salem, Mohamed I. El Adawy, 2012. "Histogram correlation for video scene change detection." In *Advances in Computer Science, Engineering & Applications*, 765-773. Springer Berlin Heidelberg.

Rasheed, Zeeshan, Mubarak Shah, 2005. "Detection and representation of scenes in videos." *Multimedia*, IEEE Transactions on, 7(6): 1097-1105.

Vlachos, Theodore, 2000. "Cut detection in video sequences using phase correlation." *Signal Processing Letters*, IEEE 7(7): 173-175.

Yeung, Minerva, Boon-Lock Yeo, Bede Liu, 1998. "Segmentation of video by clustering and graph analysis." *Computer vision and image understanding*, 71(1): 94-109.

Yi, Xiaoquan, and Nam Ling, 2005. "Fast pixel-based video scene change detection." In *Circuits and Systems, ISCAS*. IEEE International Symposium on, 3443-3446. IEEE

Young, M., 1989. *The Technical Writer's Handbook*. Mill Valley, CA: University Science.

Zhai, Yun, Mubarak Shah, 2006. "Video scene segmentation using Markov chain Monte Carlo." *Multimedia*, IEEE Transactions on, 8(4): 686-697.