MULTIMEDIA CONTENT ANALYSIS: HOMEWORK #1: FALL SEMESTER 2018

Nithin Melala Eshwarappa
Student ID: 607410141
Professor. Wei-Ta Chu

Contents

1. Abstract

In the proposed work, we are working on the concept of the object detection in the video. The object detection is in terms of the shot frame detection. The shot frame detection is nothing but the detection of the new objects which are newly getting adding into image frame and existing ones getting disappeared suddenly due the sudden movement of the video recording camera. Here the background of the frames in the video gets varied every time. Due to the changing background, the sudden appearance of the new objects in the frame which are similar to the objects as in the previous frames, the detection of the new ones together with old one is called as shot frame detection. Here we are using the threshold based method of shot frame detection by making use of a threshold which is obtained by the trial and error method.

2. Introduction

Processing of video and image provides an understanding of the scene that it describe. It is an essential component of a number of technologies including video surveillance, robotics and multimedia. It represents an area of research with huge growth in the recent past. Video shot boundary detection is one of the research works in the field of video processing. The foremost requirement of any multimedia industry is video. Over the years industries has developed comprehensive and complete measures and techniques to index, store, edit, retrieve, sequence and present video material. Shot boundary detection is usually the opening step toward automatic video indexing and browsing. It is based on the recognition of visual discontinuities caused by the transitions, to segment a video stream into elementary uninterrupted content units for subsequent high-level semantic analysis. The discontinuities usually found during scene change or shot change. These discontinuities occur in form of transitions of different kinds which are categorised into two groups: abrupt (as hard cut) and gradual (dissolve, fade in, fade out, wipe). Conventionally, if there exist frames that are merged by the adjacent shots but belong to neither of them, the transition is called a gradual one; otherwise, it is called a cut.

A shot is defined as the consecutive frames from the start to the end of recording in a camera. It shows a continuous action in an image sequence. The cut boundaries show an abrupt change in image intensity or colour, while those of fades or dissolves show gradual changes between frames. The detection of the latter is more difficult than the detection of the former. The image motion due to camera or object movements makes the shot boundary detection problem more complex. There is also a slow change in intensity for the frames with image motion. These may take their changes as shot boundaries resulting in false alarms which cause degrade the precision. Twin-comparison was developed to find shot boundaries among cuts and fades/dissolves using two thresholds. Gunsel and Tekalp proposed one threshold method using Otsu method to find the threshold automatically. However, this system was presented for detection of cut-type shot boundaries. In model-based method the edit effect showing gradual changes(fades, dissolves, etc) presents edit invariant property that is used in classifying shot boundaries. This paper proposes a framework in which accentuates edit constancy effects by applying low pass filtering to histogram differences between frames, while suppressing motion effects causing false alarms. Edit constancy effects are rectangular shapes of cut and triangular shapes of fades/dissolves in filtered histogram differences after applying window convolution to original histogram differences. And this paper also presents the shot representation method showing the key frames effectively based on contents in each shot. The most common method to select key frames is the temporal sampling method. But this method does not provide the successful representation in general. This paper presents in details the shot detection method in section 2 and the key -frame selection method.

Image analysis is the quantitative or qualitative characterization of two-dimensional (2D) or three dimensional (3D) digital images to extract meaningful information. The characterization of an image is based upon visual features which are extracted from the image. This can then be used to classify images with similar characteristics for applications such as content based image retrieval (CBIR), which is also known as query by image content (QBIC). Applications may require the classification and retrieval of the entire image as a whole; however images may also be segmented into sub-regions which represent distinct objects within the image.

3. Survey of the Existing mythologies

1) Object detection in the case of constant background

Motion detection plays a fundamental role in any object tracking or video surveillance algorithm, to the extent that nearly all such algorithms start with motion detection. Actually, the reliability with which potential foreground objects in movement can be identified, directly impacts on the efficiency and performance level achievable by subsequent processing stages of tracking and/or object recognition. However, detecting regions of change in images of the same scene is not a straightforward task since it does not only depend on the features of the foreground elements, but also on the characteristics of the background such as, for instance, the presence of vacillating elements. So, in this chapter, we have focused on the motion detection problem in the basic case, i.e., when all background elements are motionless. The goal is to solve different issues referred to the use of different imaging sensors, the adaptation to different environments, different motion speed, the shape changes of the targets, or some uncontrolled dynamic factors such as, for instance, gradual/sudden illumination changes. So, first, a brief overview of previous related approaches is presented by analyzing factors which can make the system fail. Then, we propose a motion segmentation algorithm that successfully deals with all the arisen problems. Finally, performance evaluation, analysis, and discussion are carried out.

Motion detection in scenes with background motionless, aiming at analyzing and solving different issues referred to the use of different imaging sensors, the adaptation to different environments, different motion speed, the shape changes of the targets, or some uncontrolled dynamic factors such as, for instance, gradual/sudden illumination changes. As a solution, a CoD techniques has been proposed. Mainly, it combines a *frame-by-frame difference* together with a *background(-frame) subtraction* with the purpose of overcoming the two well-known difference drawbacks (i.e., *ghosting* and *foreground aperture*). Moreover, on the way to autonomous, robust visual systems, it has also been necessary to study the automatic threshold estimation. For that, a *dynamic* thresholding method based on resolution distribution in an image has been presented. This technique automatically divides the captured images and sets the proper thresholding parameters for two different kinds of cameras: perspective and fisheye. So, problems such as non-uniform-distributed resolution, inadequate illumination gradient in the scene, unsuitable contrast, or the overlapping of the background and the target gray-level distributions, are overcome. In addition, some experiments over public image datasets and our own image datasets were carried out. Both quantitative and qualitative results have been provided in order to assess the Cod's performance under different conditions. As the experimental results have highlighted, the proposed approach is able to deal with different imaging devices, variable target's speeds or types of interest elements such as people or animals. Furthermore, a comparative analysis with some well-known techniques (those that have provided results on these image datasets) has demonstrated that our approach outperforms them. Finally, a time-execution analysis has been presented. In that study, two different parameters were considered: the image size and the process to update the reference frame for the *background (-frame) subtraction*. It highlights that the execution time depends on the image size, although a real-time performance is obtained for a 320× 240 image resolution. So, the proposed approach can be used for real-time robotic tasks.

2) Object detection in the case of Occlusion

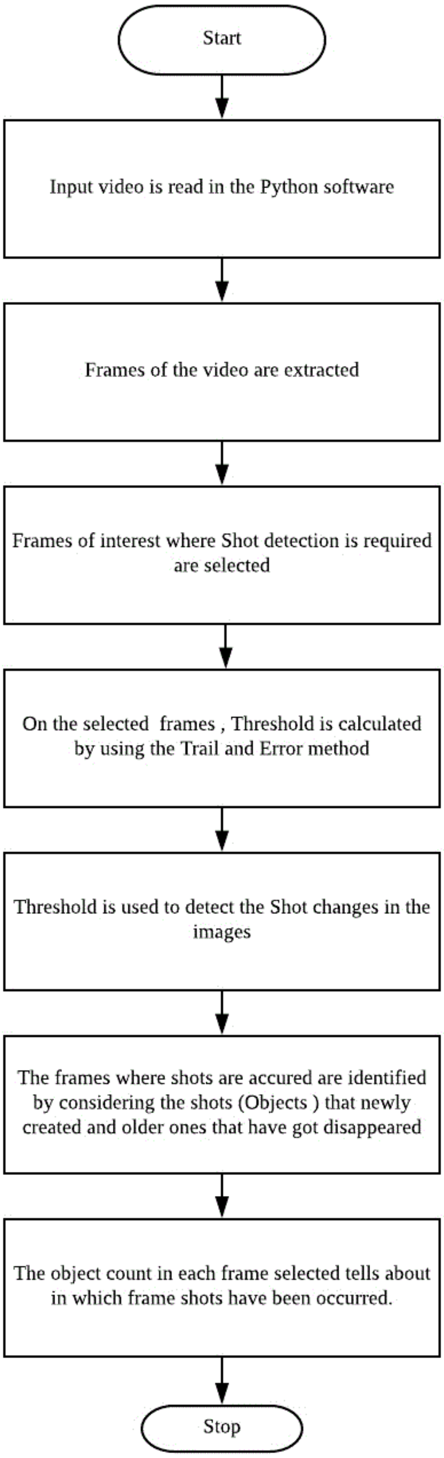
Detection of objects is a challenging task because of large variations in lighting (sun, shadows), object position and size, object deformations (shape) and large intra-class variations in object and background. Although the quality of detection algorithms is constantly improving and partially solve the previous challenges, state-of-the-art methods still struggle to detect objects that are occluded or are in unusual poses.1 Occlusion is a particular problem that is different from the previous challenges, since it takes away partial object information. The variation and amount of occlusion forms a problem of its own, which has not been broadly studied, so that we specifically investigate the handling of occlusions in this paper. Some typical occlusions are visualized in Figure 1. Popular object detection algorithms use a sliding-window detection stage, where a sliding classification window is evaluated at different positions in the image. At each search position, the local image region is classified into object/background. To remove variations in contrast and light conditions, the raw intensity values of the image pixels are typically first transformed into an invariant feature space. A popular feature descriptor for object characterization is the Histograms of Oriented Gradients (HOG).2 The obtained feature description is then classified by a linear Support Vector Machines (SVM),3 which is selected for its simplicity and good performance. A well-known dataset for occlusion experiments is the Caltech Pedestrian dataset, which focuses on pedestrian detection in an urban environment. Here, over 70% of all pedestrians are occluded in at least a single video frame. Statistics on these occlusions show that 95% of all occlusions in this dataset occur from the bottom, the right and the left of the pedestrians.4, 5 This aspect will be specifically addressed later in this paper. Our work concentrates on improving an existing real-time sliding-window object detection system that uses linear classification (SVM). To this end, we explore the detection of occluded regions and compare this with the detection of regions without occlusion. The first approach focuses on the detection of occlusions using the classification score, whereas the second approach focuses on the detection of non-occluded regions using multiple classifiers in parallel, each dealing with different partial occlusions. We will evaluate both approaches, and show that the latter approach is most suited.

3) Object detection in the case of moving background back ground

Digital image processing is among the most researched fields nowadays. The growing need of surveillance systems has further on increased the importance of this field. Surveillance systems have many applications in the field of security, intelligence gathering, traffic control and people tracking. Object detection and tracking is one of the main steps in these systems. Availability of high definition videos, fast processing computers and exponentially increasing demand for highly reliable automated video analysis has created a new and great deal for object detection and tracking algorithms [1]. The detection and tracking of the Object Of Interest (OOI) in a video or video sequence is a very important task in the automated video surveillance; In order to monitor the happenings in a particular area, to detect and recognize moving objects and to detect unlikely events and to report suspicious, criminal activities applications are developed in computer vision system. Some of the other major applications of object tracking are: Robot Vision: for navigation of robot, different obstacles in the path are identified by the steering system in• order to avoid collision. In case the obstacles are itself in motion then we need a real time object tracking system. Monitoring of Traffic: Highway traffic can be continuously monitored using moving cameras.

4. Proposed method :

The proposed method of the shot frame detection is shown in the following figure(1).



Figure(1): Proposed method of the shot frame detection

5. Steps to Run the program

- Down load the zip file names Multimedia Analytic(NIthin) zip into a folder.
- Extract the file
- Open the extracted Folder
- Open the file ‘Soccer_vedio’, soccer.py file. In Edit with IDLE 3.6(64bit) tool of Python Software.
- Run the Soccer.py file using F5 Function Key.

6. Result:

Figure(2): Original shot frame detection

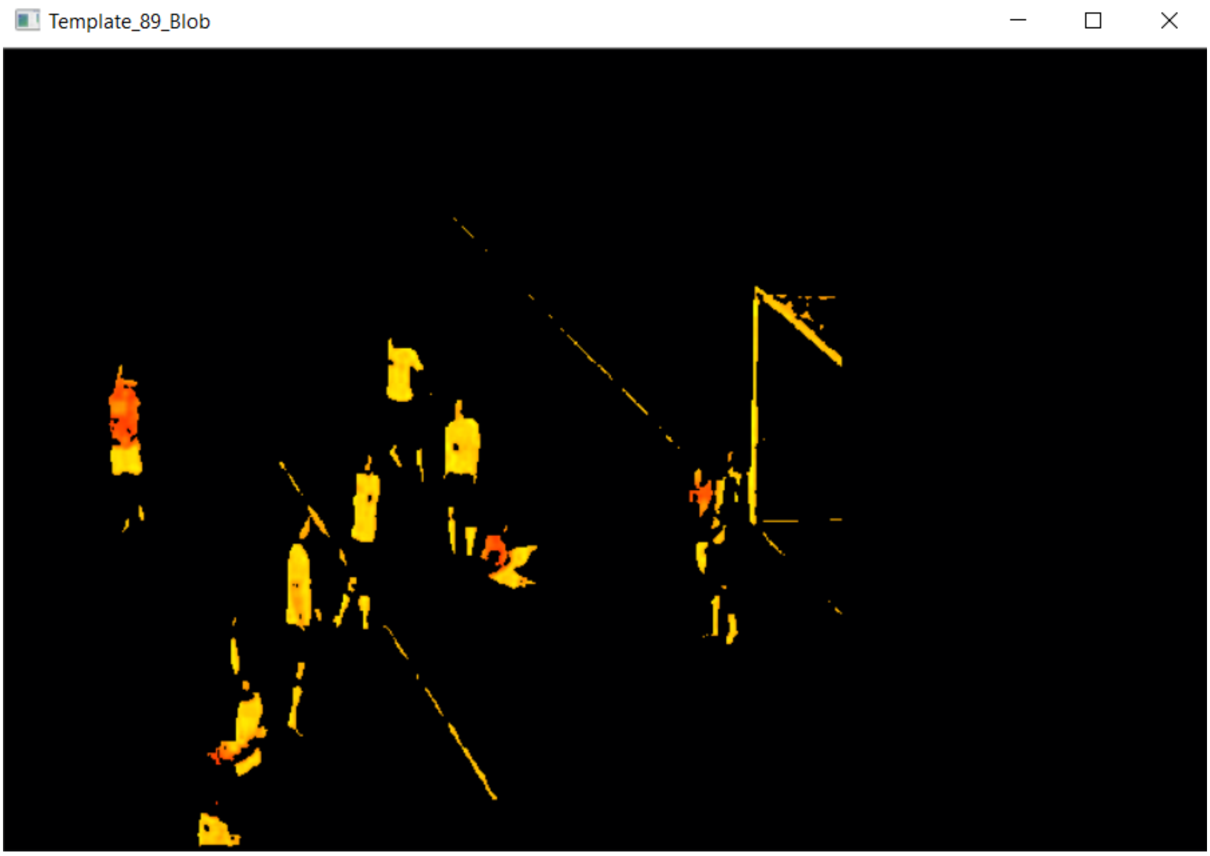


Figure (3): Template frame of the shot frame detection

7. **Receiver Operating Characteristic curve**

- This type of graph is called a **Receiver Operating Characteristic curve** (or ROC curve.) It is a plot of the true positive rate against the false positive rate for the different possible cut points of a diagnostic test.
- It shows the trade off between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity).
- The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test.
- The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.
- The area under the curve is a measure of text accuracy

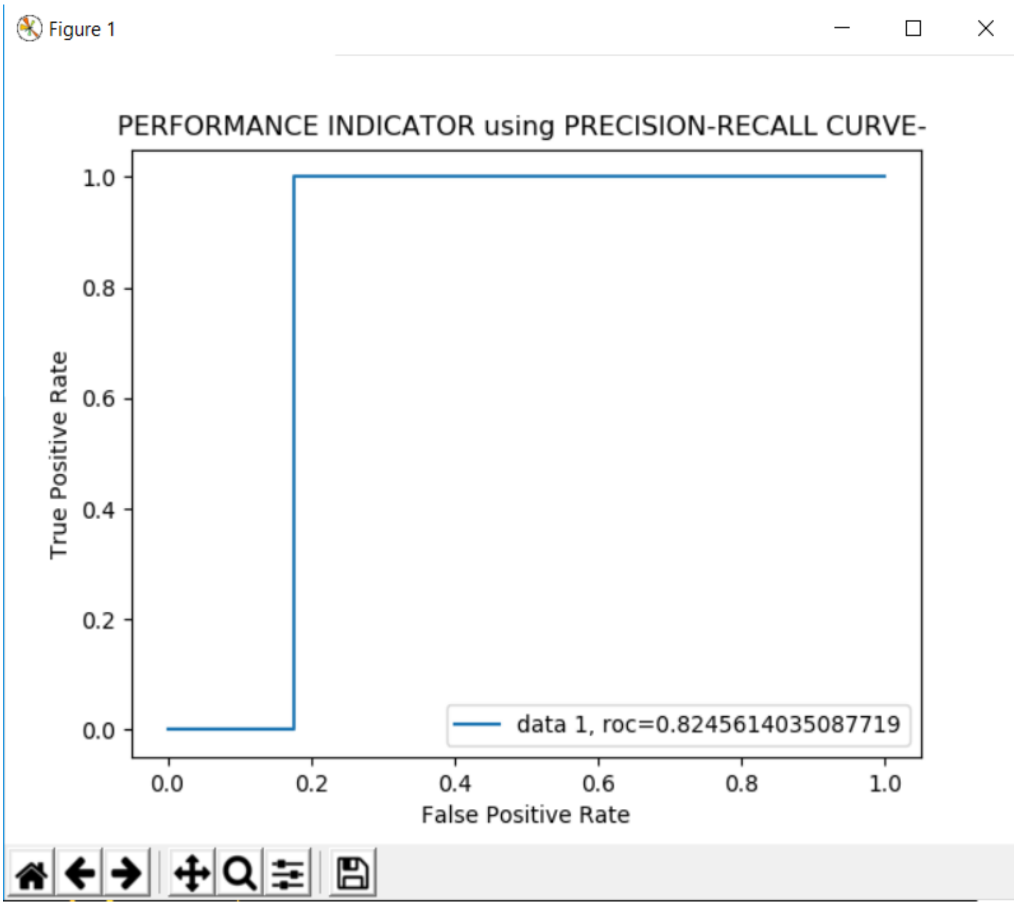


Figure (4): Receiver Operating Characteristic curve

Conclusion

Here in the given video, the frames are extracted and the few frames where the shots frames are observed are identified and these identified frames are selected. In these identified frames, the objects are identified by using the proposed trial and error based method of object detection and shots that are occurring in the frames are identified and detected and are highlighted. The algorithm and program to identify the shot boundary has been made using Python. Some steps include in the program. The beginning of the program is initialized some variable and filename, loading movie and rendering the *.mpg file. The next step is finding shot frame detection using the trial and error method. After identifying the new frame, the program continues to calculate the Precision-Recall and Receiver Operator Characteristic Performance. From the experiments, the proposed algorithm can handle nicely though not perfectly identify the shot in the Soccer movie and News. However, the P-R and ROC badly achieves the result in the situation where there are too much objects with extreme changing frame intensity in a single shot.

References:

1. Kameda, Y., Minoh, M.: A human motion estimation method using 3-successive video frames. In: International Conference on Virtual Systems and Multimedia (VSMM), pp. 135–140. Gifu, Japan (1996)

2. Kanade, T., Collins, R., Lipton, A., Burt, P.,Wixson, L.: Advances in cooperative multi-sensor video surveillance. In: Darpa Image UnderstandingWorkshop, vol. I, pp. 3–24. Morgan Kaufmann (1998)

3. Migliore, D., Matteucci, M., Naccari, M.: A revaluation of frame difference in fast and robust motion detection. In: 4th ACM International Workshop on Video Surveillance and Sensor Networks (VSSN), pp. 215–218. Santa Barbara, California (2006)

4. Wren, C., Azarbeyejani, A., Darrell, T., Pentland, A.: Pfinder: Real-time tracking of the human body. IEEE Transactions on Pattern Analysis andMachine Intelligence (PAMI) **19**(7), 780–785 (1997)

5. Sezgin, M., Sankur, B.: Survey over image thresholding techniques and quantitative performance evaluation. Journal of Electronic Imaging **13**(1), 146–168 (2004)

6. Miˇcušík, B.: Two view geometry of omnidirectional cameras. Ph.D. thesis, Center for Machine Perception, Czech Technical University in Prague (2004) 8. Toyama, K., Krumm, J., Brumitt, B., Meyers, B..