International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

A New Method for Video Shot Detection Using Visual Bag Using Corner Detector

Vindhya G. B[1], Dr. Jharna Majumdar[2]

[1]MTech, 4th Semester, Computer Science and Engineering, Nitte Meenakshi Institute of Technology, Bangalore, India
[2]Dean R&D, Professor, Head of Department, M.Tech, Computer Science & Engineering, Nitte Meenakshi Institute of Technology, Bangalore, India

Abstract: ***In this paper, a system called Video Shot Detection using Visual Bag of Words is proposed for effective shot change detection using multiple feature extraction algorithms with the help of corner detectors. These multiple features which form Bag of Visual Words are extracted in order to retrieve patches from each frame. The detection method was tested on different videos such as sports, entertainment and other documentary videos which contain different types of shots having different objects. Our results have shown that the system can produce accurate shot change detection which is useful for video content analysis and indexing.***

Keywords: Feature Extraction, Bag of Visual Words

1. Introduction

Video Shot Detection is also called "Cut Detection", is a field of research of video processing. The objective of cut detection is finding the position in the video in that one scene is replaced by another one with different visual content. It is fundamental to any kind of video analysis and video application since it enables segmentation of a video into its basic components: the shots. A video shot is defined as a continuously imaged temporal segment of a video. Since the semantic content of a video is largely based on its production (imaging) process, the complete segmentation of a video into shots is a fundamental first step for most kinds of semantic video processing tasks. Classification of shot boundary detection algorithms, including those that deal with gradual shot transitions. For lack of space we give neither a general introduction to the shot change detection problem, nor a detailed look at specific algorithms. Applications of Video Shot Detection are to automate the indexing and retrieval and management of videos. Structural analysis of video is the basic step in analyzing video content developing techniques for efficient access, classification, retrieval, and browsing of large video databases. Amongst the various structural levels (i.e., frame, shot, scene, etc.), shot level organization has been regarded suitable for browsing and content-based retrieval. Another popular approach has been proposed for Content based image retrieval which is called Bag of Visual Words.

Bag of Visual Words is defined as model which is a simplifying representation used in natural language processing and information retrieval, whereas Visual Words is a small part of an image(PATCH) which will carry some kind of information related to features. Ex: Color, Shape, Texture, etc. Thus Bag of Visual Words (BoVW) is defined as a model which is a collection of visual words which together can give information about the meaning of image at all. Initially BoVW model was proposed for large scale image retrieval with the help of content based methods in order to retrieve the semantic content of image, which in turn grasps the semantic similarity between the content between images directly. Thus we proposed a model where we adopted this BoVW model method to classify the shots by retrieving the semantic content in the form of PATCHES from each frame. In order to retrieve the PATCHES, features are applied for each frame which in turn obtains the PATCH. So to convert the content of a frame into patches, initially Corner Detectors are applied which will give Key points or Interest points in an image to which features are applied in order to form a PATCH.

A Corner point is basically an Interest point as they are formed from two or more edges which usually defines the boundary between two different objects or parts of the same object. Many Corner detectors have been developed but in our proposed model, we applied Moravec and Harris Plessey Corner Detector.

2. Proposed Model

Algorithm should be designed with the reduction of number of false detections in the presence of object and free from the complexity of computational time in mind. Here, we propose a Bag of Visual Words model which will extract set of features from each frame in order to convert each frame into **PATCH** and compare the corresponding features between the successive frames to measure the visual content similarity between two frames. The flowchart for proposed model is as shown in **Figure 1**.

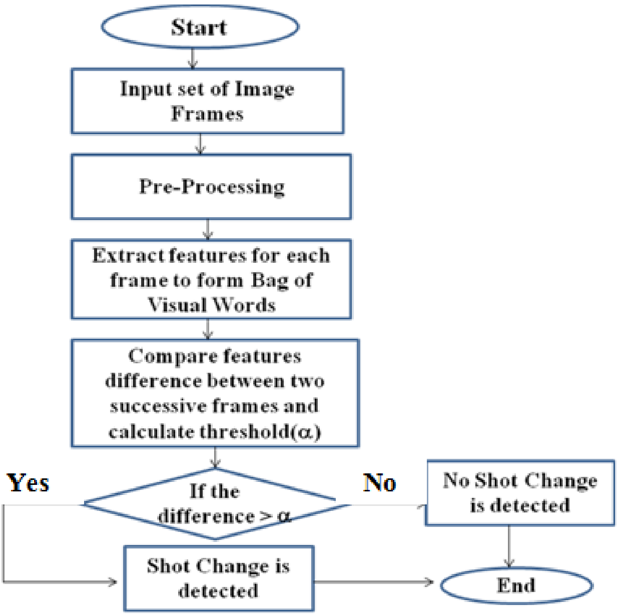

Figure 1: Flowchart

2.1Algorithm for Feature Extraction

Figure 2 explains the proposed system. The flowchart defines the algorithm for feature extraction. Before explaining the algorithm in brief let us know the Pre-processing step which is shown in the figure 2 which is to be done before features are extracted.

Step 1: Pre-Processing

(a)

i. For each image frame, apply moravec corner detector.
ii. Use the threshold value to limit the number of corner points.

(b)

i. Divide the image fame into number of equal sized blocks.
ii. For each block, find the number of corner points in each block.
iii. Sort them in descending order.
iv. Take the first three blocks having maximum corner points.

Step 2: Feature Extraction

(a) To convert these blocks into patches, features are extracted for the corner points present in each of these block.

i. These features include

Feature 1: Left Profiling,
Feature 2: Right Profiling,
Feature 3: Top Profiling and
Feature 4: Bottom Profiling.
Feature 5: Centroid
Feature 6: Polygon

(b) Profiling:

The profiling feature is shown in the **Figure 2**.

ii. i. Left Profiling: Distance from left side of the image block to corner points.
iii. ii. Right Profiling: Distance from right side of the image block to corner points.
iv. iii. Top Profiling: Distance from top of the image block to corner points.
v. iv. Bottom Profiling: Distance from bottom of the image block to corner points.
vi. v. Features should be made invariant and must be stored in an array which forms four features for each patch.
vii. Repeat this for all patches.
viii. These features are compared with the successive frames and a shot change is detected.
ix. For Shot Change Detection:

a. Horizontal Ratio:

(Summation of left distance)/(Summation of right distance)

b. Vertical Ratio:

(Summation of top distance)/(Summation of bottom distance)

xi. Difference between two frames:

a. Euclidian distance = Sqrt(((Horizontal Ratio1-Horizontal Ratio2)*(Horizontal Ratio1-Horizontal Ratio2)) + ((Vertical Ratio1-Vertical Ratio2)*(Vertical Ratio1-Vertical Ratio2))
x. If Euclidian distance > □ = Shot Detected.
xi. Otherwise shot change is not detected.

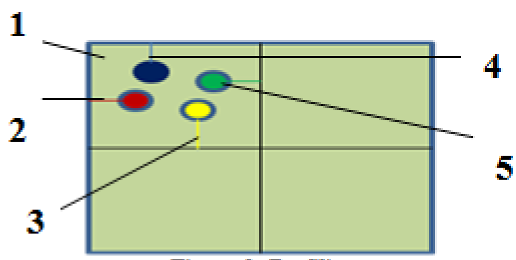

Figure 2: Profiling

Where,

1 ⟶ Image Block
2 ⟶ Left Profiling
3 ⟶ Bottom Profiling
4 ⟶ Top Profiling
5 ⟶ Right Profiling

(c) Centroid

The Centroid feature is shown in **Figure 2**

i. Find the centre of image block.
ii. Find the distance from centre of image block to all the corner points.
iii. Find minimum and maximum distances form calculated distances.
iv. Find ratio of maximum distance/minimum distance.
v. For Shot Change Detection:

a. Centroid Ratio1(Frame 1):

(MaximumDistance)/(Minimum Distance)

b. Centroid Ratio2(Frame 2):

(Maximum Distance)/(Minimum Distance)

vi. Difference between the two frames:
a. Euclidian Distance = Sqrt(((Centroid Ratio1 – Centroid Ratio2)*(Centroid Ratio1 – Centroid Ratio2)))
vii. If the Euclidian Distance >□, shot change is detected.
viii. Otherwise no shot change is detected.

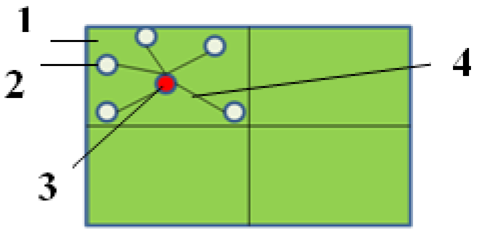

Figure 3: Centroid

Where,

1 → Image Block

2 → Corner point

3 → Centroid corner point

4 → Distance from centroid to all corner points

(d) Polygon

i. Find the corner points which has minimum distance for left, right, top and bottom line of the image block.
ii. Draw a polygon by joining all the four points.
iii. Find the inner angle of polygon.
iv. Find the ratio of : (area of inner box)/ (area of image block).
v. Find the ratio of: (maximum angle) / (minimum angle).
vi. Angle of polygon= Area of Rhombus= a*b*sin□□

a= distance between x1 and x2.
b=distance between x1 and x4.
dx1= (x2-x1), dy1= (y2-y1),
dx2= (x3-x1), dy2= (y3-y1)
m1= sqrt((dx1*dx1)+(dy1*dy1))
m2= sqrt((dx2*dx2)+(dy2*dy2))
□=acos(dx1*dx2+dy1*dy2)/(m1*m2)*(180/ п)

vii. Area of polygon= (0.5*b*h),
b= distance between x1 and x3,
h= distance between x2 and x4.
Area of Image block = Block Width * Block Height

viii. For shot change detection:
a. Take the difference of Euclidian distance between angle and area ratios of two frames as specified in step 3 in second set of features.
xi. If Euclidian distance >□, shot change is detected.
xii. Otherwise shot change is not detected.

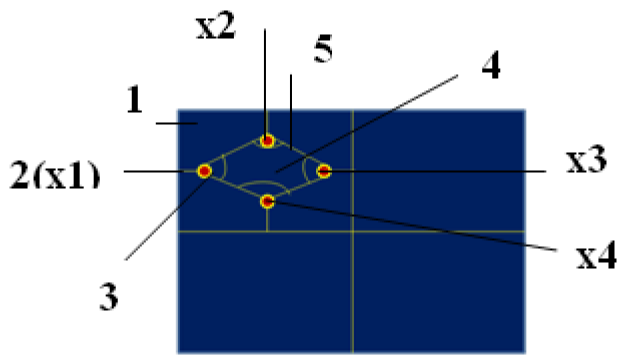

Figure 4: Angle and Area of Polygon

Where,

1 → Area of Image Block

2 → Corner points with minimum distance from all the four sides.

3 → Angle of the polygon (θ)

4 → Area of polygon

5 → Polygon

3. Results

i. All algorithms discussed above are implemented on four different types of videos such as Football.avi, Hurricane.avi, Cut.avi and Movie.avi
ii. But the results are shown for only Hurricane.avi video in Figure 4, 5, 6(□ varies between 1 and 5).

Figure 4: Shots detected for Profiling

Figure 5: Shots detected for Centroid

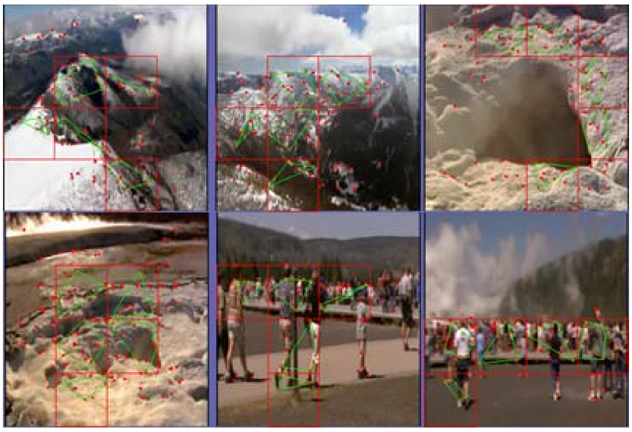

Figure 6: Shots detected for Polygon

3.1 Observation and Analysis

The result obtained for all the three algorithms shown has to be analyzed to measure the performance of all the three algorithms. To measure this performance, accuracy has to be calculated. Accuracy is measured with the help of Precision and Recall which is done for four different types of videos such as Football.avi, Hurricane.avi, Cut.avi and Park.avi.

- Recall: (Correct)/(Correct + Miss)
- Precision: (Correct)/(Correct + False)

Table 1

Input Video	*No. of Frames*	*Shots detected*	*Precision*	*Recall*
Football.avi	1711	9	75%	100%
Hurricane.avi	1390	5	83%	85%
Movie.avi	617	5	62%	72%
Cut.avi	610	6	85%	89%

Table 2

Input Video	*No. of Frames*	*Shots detected*	*Precision*	*Recall*
Football.avi	1711	9	75%	95%
Hurricane.avi	1390	5	71%	91%
Movie.avi	617	5	83%	82%
Cut.avi	610	6	75%	100%

Table 3

Input Video	*No. of Frames*	*Shots detected*	*Precision*	*Recall*
Football.avi	1711	9	81%	90%
Hurricane.avi	1390	5	62.5%	85%
Movie.avi	617	5	55%	79%
Cut.avi	610	6	60%	80%

Form the above observation we can analyze that Centroid method performs better than the rest of two methods and also among the two corner detectors, Harris works best.

4. Conclusion

Visual Bag of Words has provided an overview of a model which was first proposed for text retrieval system and then due to further enhancements it is applicable in many other applications such as Video Shot Detection, Robotic applications, etc. So, it is also proved that Visual bag of Words has performed well in the respective fields specified above.Thus we adopted this popular approach for object class detection. To the authors' best knowledge, our work is the first which uses BoVW approach which will classify the shots based on the features extracted from each frame. We utilized the efficient implementations in our method. In future work, we will investigate other video processing tasks using BoVW approach and optimization of our implementation to run with low computation time.

References

[1] Jialu liu, "Image retrieval based on Bag of Words Model," 2009.
[2] Maja Rudinac, Boris Lensiegne, Peiter Jonker, "Key point extraction and selection for object recognition", 2009.
[3] Ravi Shekar and C.V Jawahar, "Word Image Retrieval Using Bag of Words," 2012.
[4] Mohammed Ally, Mario Munich, Pietro Perona, "Multiple Dictionaries for Bag of Words Large scale image retrieval", 2010.
[5] Jun Yang, Yu-gang Jiang, Alexander G Hauptmann, Chong-Wah Ngo., "Evaluating Bag of Visual Words in Scene Classification," 2007.
[6] Jukka Lankein,Joni-Kristian," Video Shot Detection Using Visual Bag of Words", 2008

[7] Kristen Grauman, Bastian Liebe, “Indexing and Visual Vocabularies”,2009
[8] Chris Harris and Mike Stephens, “A Combined Corner and Edge Detector, 1988.
[9] Jing Wang, JingDong Wang, Qifa Ke§ Gang Zeng, Shipeng Li, “Fast Approximate Kmeans vis Cluster Closures”, 2006.
[10] Gita Sinha, Mrs. Rajneesh Rani, Prof. Renu Dhir, “International Journal of Advanced Research in Computer Science and Engineering”, 2012.
[11] Yang Mingqiang, Kpalma Kidiyo, Ronsin Joseph, “A Survey of Shape Feature Extractions”, 2008
[12] “Handwritten Character Recognition Using Multiclass SVM Classification with Hybrid feature extraction”, 2012