

EVENT RECOGNITION: IMAGE & VIDEO SEGMENTATION

Pavan Kumar H J¹, Mangala C N²

¹Student, Department of CSE, East West Institute of Technology, Bangalore, Karnataka, India

²Asst Professor, Department of CSE, East West Institute of Technology, Bangalore, Karnataka, India

Abstract

This paper gives a clear look at the segmentation process at the basic level. Segmentation is done at multiple levels so that we will get different results. Segmentation of relative motion descriptors gives a clear picture about the segmentation done for a given input video. Relative motion computation and histograms incrementation are used to evaluate this approach. Also here we will give complete information about the related research which is done about how segmentation can be done for the both images and videos.

Keywords: Image Segmentation, Video Segmentation.

1. INTRODUCTION

Images are used to represent objects, living things and things which are present around us. Image can be taken by the camera or can be viewed through eyes. The continuous movement of images is called video. Video also can be captured by camera or can be watched through eyes. Any image or video is not 100% perfect. So here we try to emphasize our work so that clarity of that video should be higher.

Previously, patches in the given frame alone are used to bring back the original image. The factors such as with corresponds to the images such as relative motion of the object, noise in a given image and the camera angle were played a vital role in identifying the original image. Pair-wise relationships are used to extract image from the video. Dense sampling is used to extract frames from videos.

As with the earlier systems, patches alone are not efficient to bring back the original image. The various factors relative motion of the object, noise present in the image, camera angle are less robust to the methods such as tracks or tracklets which were used to take recover the patched image. Pair-wise relationships are used to extract image from the video, but this method fails, because after employing this method video will get fragmented. Dense sampling is one of the recent methods used to develop the patched image; however it's not feasible to develop for large datasets

2 LITERATURE SURVEY

Paper1: Evaluation of Local Spatio-Temporal Features for Action Recognition

This paper mainly concentrates on the recognizing of actions in given videos by using local spatio-temporal feature. Here they developed a common evaluation setup by using the datasets KTH, UCF sports and Hollywood2 which are old.

Disadvantages

- Hollywood2 datasets fails to detect the cureent feature of the given video.
- UCF and KTH are less efficient when compared to Harris3D, Cuboid and Hessain datasets.

Paper 2: Mosift Recognizing Human Actions In Surveillance Videos

The main goal of this paper is to recognize the human actions in give any type of videos. Local spatio-temporal features are used for this purpose.

Disadvantages:

- It lacks in finding the patches in either 2D/3D images.
- It won't use frames to recognize the actions. Hence output will be very low.
- Bag of words approach is not well suited.

Paper 3: Hierarchical Spatio-Temporal Context Modeling for Action Recognition.

Recognizing actions in any given video will be a difficult task.

To do object detection and recognition in a given video we need some advanced and perfect techniques.

Disadvantages:

- Complete evaluation of the image is not done here.
- Relative motions are not examined.
- It fails to explain why it's not possible to select all the important pixels.

Paper 4: A SURF-based Spatio-Temporal Feature for Feature-fusion-based Action Recognition

Here methods used such as SURF-based light-weight spatio-temporal feature which is good for large-scale video

datasets. Maintaining data sets will be a difficult job. Efficiency of KTH dataset and Youtube data set is around less than 90%.

Disadvantages:

- Wastage of database if we used a small number of objects. Because here we will maintain a large database.
- Efficiency should be more.
- Fails to take the image or video in 360° angle.

Paper 5: Action Recognition by Dense Trajectories

Here we used dense sampling along with the feature tracking for analyzing the videos. Here the image/video captured is relatively having a better quality.

Disadvantages:

- Usage of multiple datasets increases the overhead on the database
- Efficiency between the datasets is not consistent.

3. PROBLEM STATEMENT

Patches in a given frame from a video is recovered with the help of the other relative frames.

4 PROPOSED SYSTEM

In proposed system we mainly concentrate on the relative motion of the object, noise in a given image and the camera angle. Along with the patched images, relative motion frames are utilized to get back the patched image. Also segmentation can be done to more than one type of videos(.avi, .wmv).

5. ALGORITHM

5.1 Relative Motion Histogram Computation.

The motive of algorithm is used to convert the image(.jpeg)/ video(.avi/.wmv) into a segmented image/video.

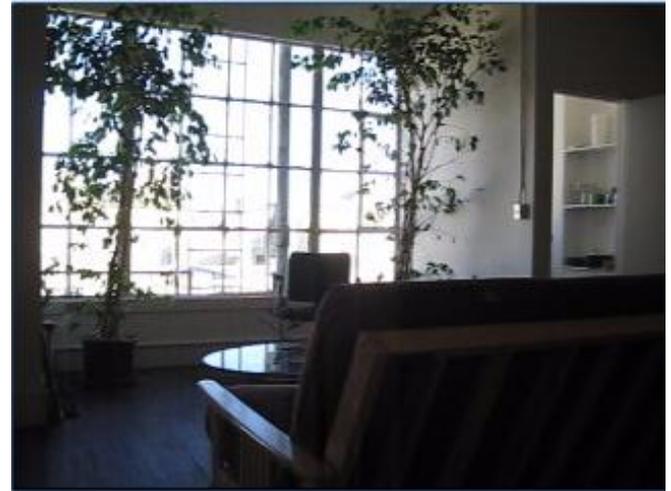
With the help of this algorithm we will convert an image from a given .avi or .wmv video to segmented image.

The steps for the above algorithm as follows:

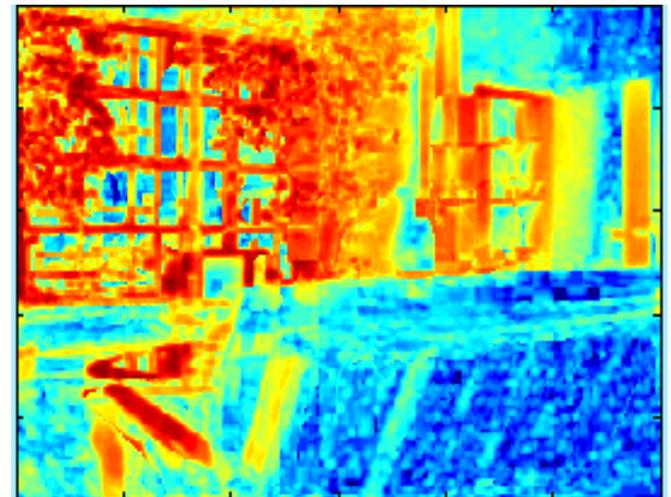
- Superpixel comparison: Comparing the boundary of superpixels of the analyzed volume with all the superpixels of other volumes within a defined spatio-temporal neighborhood.
- Relative motion computation associating the neighboring volume motion to its underlying superpixels.
- Quantization according to temporal distance, spatial distance and direction.
- Intensity and histograms are implemented on the basis of quantization.

5.2 Snapshots

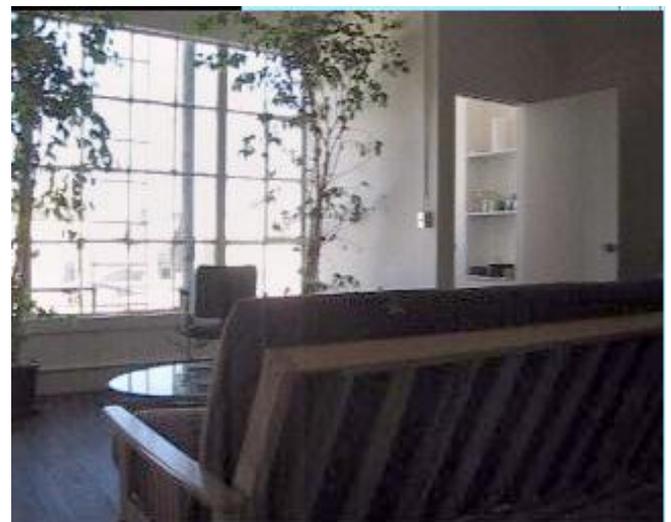
Original Image



Segmented Image



Output Image



6. ADVANTAGES OF PROPOSED SYSTEM

By using relative motion frames along with the patched images we will get a better quality for the patched image. Ground video motion and appearance descriptors on a video segmentation producing a hierarchical, multi-scale output. We can take out patched image from given any type of video.

7. CONCLUSION

With the above algorithm I can get a segmented image in less time compared to all other segmentation algorithms. Pixel comparison is one of the best methods to get/match the pixels from the original image. To get differentiate between the neighboring pixels neighboring volume motion is used. Color assigned to each pixel on the basis of quantization. Then the results are drawn with the help of histograms.

As with the future enhancement, the original video retraced back so that the enhanced image/video is having more clarity

REFERENCES

- [1]. H. Wang, M. M. Ullah, A. Klaser, I. Laptev, C. Schmid, *Evaluation of local spatio-temporal features for action recognition*, BMVC 2009.
- [2]. M.-Y. Chen and A. Hauptmann, *MoSIFT: Recognizing Human Actions in Surveillance Videos*, CMU-CS-09-161, Carnegie Mellon University, 2009.
- [3]. J. Sun, X. Wu, S. Yan, L.-F. Cheong, T.S. Chua, and J. Li, *Hierarchical spatio-temporal context modeling for action recognition*, CVPR, 2009.
- [4]. A. Noguchi and K. Yanai, *A SURF-Based Spatio-Temporal Feature for Feature-Fusion-Based Action Recognition*, ECCV, 2010.
- [5] H. Wang, A. Klaser, C. Schmid, and L. Cheng-Lin, *Action Recognition by Dense Trajectories*, CVPR, 2011.

BIOGRAPHIES



PAVAN KUMAR H J received Bachelor of Degree from Bengaluru College of Engineering & Technology, Bengaluru. Now doing pursuing M. Tech in East West Institute of Technology, Bengaluru.



Smt. MANGALA C N Asst. Professor in East West Institute of Technology, Bengaluru. Her area of specializations is mainly of Image processing.