INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

SURVEY ON CLASSIFYING ENERGY FEATURES FOR VIDEO SEGMENTATION

J. Anusha

Department of CSE & Karunya University

anushajegadeesan1991@gmail.com

Abstract
Video segmentation is a major role in digital image processing. This paper is based on the survey of various video segmentation techniques to extract region of interest which is coherently moving object. Support Vector Machine Classifier is used to classifying the energy feature. Each technique has its own advantages as well as disadvantages.

Keywords: — Video segmentation, energy features, Support vector machine, coherent motion, saliency

1. INTRODUCTION

Image processing is any form of signal processing for which the input is an image. The result of image processing is an image or a set of characteristics. Most image-processing techniques occupy treating the image as a two-dimensional image and applying standard signal-processing techniques to it. Image is composed of a finite number of elements called pixel. Different types of images are binary images, intensity images, indexed images and RGB images.

Video segmentation is the process of dividing a sequence of frames into smaller meaningful pixels. This process serves as a primary step towards any additional analysis on video frames for content analysis. One of the applications for segmentation is separation of foreground (object) and background (noise). Video segmentation is mainly used for extracting region of interest which is coherently moving object. Region of interest is calculated by

classifying the energy features. Classification is done by the use of Support vector machine classifier. Dynamic moving objects are detected and classified according to the energy features. Classification and pattern recognition can be done by using Support Vector Machine Classifier. In machine learning, support vector machines are supervised learning models with associated learning algorithms that examine data and recognize patterns. The SVM takes a set of input data and predicts which of two possible classes forms the output

2. METHODOLOGY

2.1 Spatio-temporal saliency model

Spatio-temporal saliency model is mainly used to predicting the eye movement during video viewing. When people viewing video, their concentration on some salient region. This salient region attracts attention of the people. Consider video is the input for this model. From each frame, Retina model extracts two signals. Each signal is decomposed into features by cortical-like filters. These filters are used to extract static and dynamic information. According to frequency selectivity and two saliency maps (static and dynamic), static and dynamic information's are extracted. Finally spatio-temporal saliency map is obtained by combining both saliency maps per video frame. This map used to predicts the gaze direction to particular areas of the frame analysed.

In Retina model, frame information is flows to photoreceptors. Photosensitive cell in the retina is photoreceptors. Then the information goes from the photoreceptors to the horizontal cells. The bipolar cells take the difference of the outputs of the photoreceptors and the horizontal cells. Amacrine cells give a second local average of the bipolar cells output [1]. Retina has two outputs, there are parvocellular and magnocellular output. Parvocellular output enhance frame contrast, which attracts human gaze in static frame.

Then the visual information is decomposed into spatial frequencies, orientations, colours and motion. Cortical-like filters used to decomposing the signal into features. This model uses Gabor filter, which is used to model V1 cells to extract frequencies, orientations and motion information. Static and dynamic information's are extracted by the Gabor filter. Dynamic saliency is related to motion and motion of a region. Speed of moving region against background was computed using motion estimator on compensated frames at the Magnocellular output of the retina. If a pixel had a motion in one frame but not in the previous frame, it is noise resulting from the motion estimation. To remove noise from the frames, temporal median filter was applied. The static and dynamic maps are fused to create the spatio-temporal saliency map [1]. This map is mainly used to predict the areas that would be gazed at by people when looking at the videos.

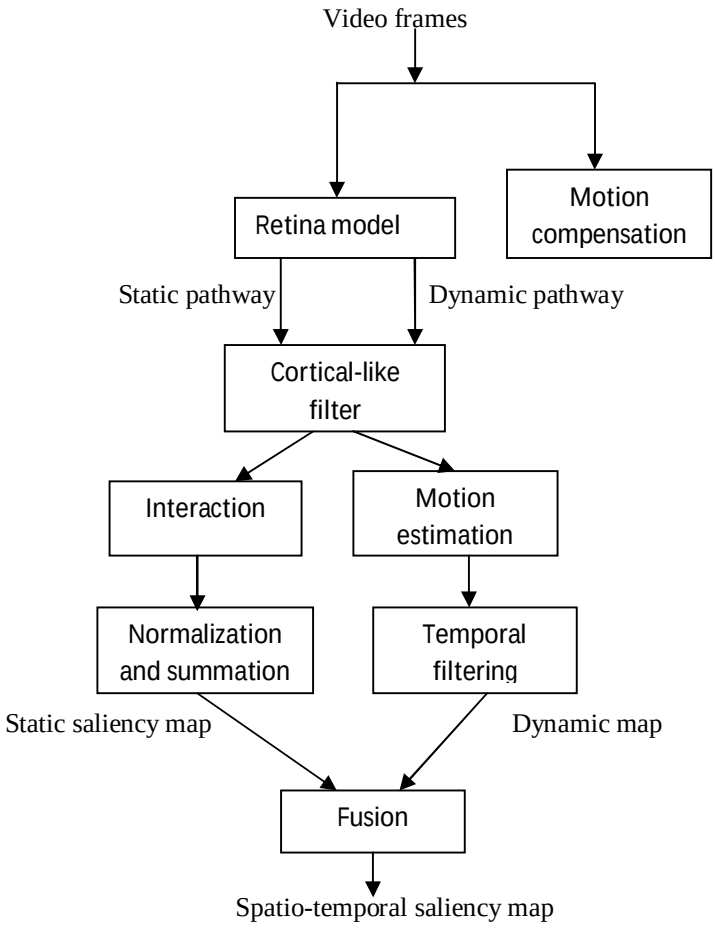

Fig. 1 Spatio-temporal saliency model

2.2 Motion saliency maps

The human visual system is able to detect coherent motion and select within multiple moving objects the most conspicuous or most relevant to the task at hand. Artificial agent operating in dynamic environments needs to be endowed with a mechanism for rapid detection and prioritization of moving stimuli in its field of view. Biologically and psychologically inspired model of this ability and tune it for the extraction of motion at different scales and velocities. Many computational models calculates saliency pixel wise and extract moving proto-objects through segmentation of motion energy features. These perceptual units called proto-objects; these are identified as consistently moving blobs.

A proto-object based priority map is obtained by assigning a single saliency value to the region confining a segmented object. Priority root from a combination of bottom up saliency is evaluated in a centre surround fashion and from top down biasing of motion features or motion saliency. This model builds upon a novel approach for extracting and prioritizing moving objects in a scene [6]. This is based on low-level processing and relies on the extraction of coherent motion in different directions. Higher-level processing has to be combined with specific task

descriptions and a more elaborated description of motion patterns in terms of frequency and spatiotemporal signatures. Extracted proto-object patches defined as blobs of consistent motion in terms of module and direction. Gestalt law of common fate states is points moving with similar velocity and direction are perceptually grouped together in a single object.

The following figure represents the flow for motion extraction and saliency computation. In the beginning a frame buffer is filtered by a Gabor filter bank and direction based feature maps are obtained (R, L, U, D). Afterwards, horizontal and vertical components of motion energy are computed ( ) and from those energy magnitude and phase are extracted, which allows the segmentation of proto-objects upon which priority is finally computed

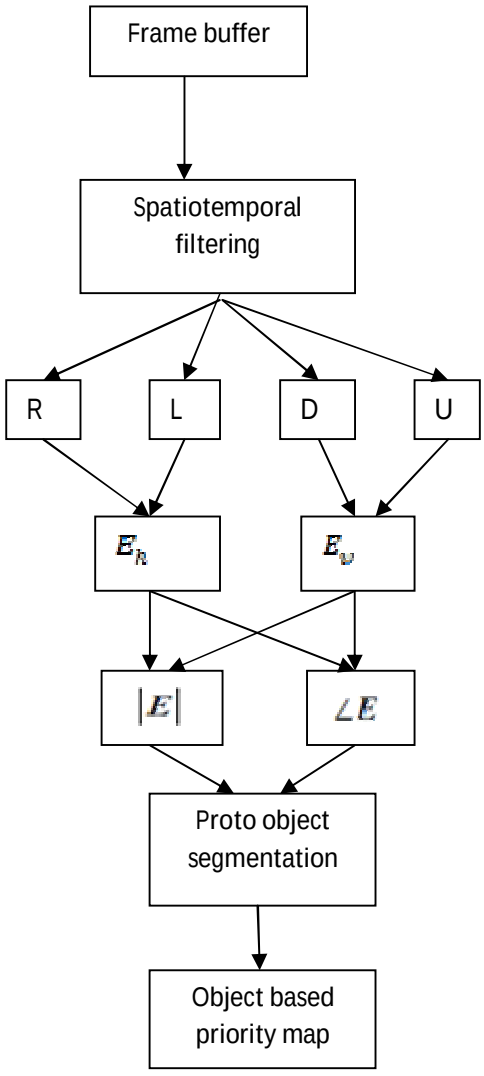

Fig. 2 Flow diagram for motion extraction and saliency computation.

2.3 LIBSVM

LIBSVM is a library for Support Vector Machines (SVMs). The goal is to help users to easily apply SVM to their applications. LIBSVM gained wide attractiveness in machine learning and many other areas [2]. Presented issues are solving SVM optimization problems, probability estimates, theoretical convergence, multi-class classification and parameter selection. LIBSVM supports SVM formulations for classification, regression and distribution estimation.

To train SVM problems, users must specify some parameters. LIBSVM provides a simple tool to check a grid of parameters. For each parameter setting, LIBSVM obtains cross-validation (CV) accuracy. Finally, the parameters with the highest CV accuracy are returned. The parameter selection tool assumes that the RBF (Gaussian) kernel is used although extensions to other kernels and SVR can be easily made. Algorithms used in LIBSVM to resolve dual quadratic problems. The first part considers optimization problems with one linear constraint, while the second part checks those with two linear constraints.

2.4 Attention based Information maximization (AIM)

Visual saliency computation built on a first principles information theoretic formulation dubbed Attention based on Information Maximization (AIM). This comprises a principled explanation for behavioural manifestations of AIM. AIM is built totally on computational constraints and the resulting model structure exhibits considerable agreement with the organization of the human visual system. Density Estimation is producing a set of independent coefficients for every local neighbourhood within the image yields a distribution of values for any single coefficient based on a probability density estimate in the form of a histogram or Kernel density estimate. Combined Likelihood is coefficient may be readily converted to a probability by looking up its likelihood from the corresponding coefficient probability distribution derived from the surround.

The product of all the individual likelihoods corresponding to a particular local region yields the joint likelihood. The joint likelihood is translated into Shannon's measure of Self-Information. The resulting information map depicts the Saliency attributed to each spatial location.

2.5 Theory of Visual Attention

The basic theory (TVA) combines the biased-choice model for single-stimulus detection with the fixed-capacity independent race model (FIRM) for selection from multi-element displays. TVA organizes a large body of experimental findings on presentation in visual recognition and attention tasks. A recent development (CTVA) combines TVA with a theory of perceptual grouping by proximity [4]. CTVA explains effects of perceptual grouping and spatial distance between items in multi-element displays. A new account of spatial focusing is

proposed in this paper. The account provides a framework for understanding visual search as interplay between serial and parallel processes.

Importantly, static and dynamic processing streams are merged at the level of visual proto-objects, that is, ellipsoidal visual units that have the additional medium-level features of position, size, shape and orientation of the principal axis. Proto-objects serves as input to the TVA process that combines top-down and bottom-up information for computing attentional priorities so that relatively complex search tasks can be implemented. To this end, separately computed static and dynamic proto-objects are filtered and subsequently combined into one combined map of proto-objects. For each proto-object, attentional priorities in the form of attentional weights are computed according to TVA. The goal of the next saccade is the centre of gravity of the proto-object with the highest weight according to the task.

Above methods have some disadvantages. There are optimization problem, high computation efficiency and difficult to determine how many subjects looking at the same location. To overcome these problems, Support vector machine classifier is involved which is used to correctly classifying the feature for extracting or determining object presented in the video.

3. CONCLUSION

Video segmentation is one of the object detection processes that can be used in the image processing. Segmentation process can be done by the use of energy feature. There are various techniques that are used for the effective and accurate object detection and extraction. In this paper various techniques are analysed. And also explains various advantages and disadvantages of various techniques.

REFERENCES

[1] Mital, P.K., Smith, T.J., Hill, R.L., Henderson, J.M., 2011. Clustering of gaze during dynamic scene viewing is predicted by motion. Cognit. Comput. 3 (1), 5–24.

[2] Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines.

[3] Mahadevan, V., Vasconcelos, N., 2009. Spatiotemporal saliency in dynamic scenes. IEEE Trans. Pattern Anal. Machine Intell. 32, 171–177.

[4] Wischnewski, M., Belardinelli, A., Schneider, W.X., Steil, J.J., 2010. Where to look next? Combining static and dynamic proto-objects in a TVA-based model of visual attention. Cognit. Comput. 2 (4), 326–343.

[5] Dorr, M., Martinetz, T., Gegenfurtner, K.R., Barth, E., 2010. Variability of eye movements when viewing dynamic natural scenes.

[6] Marat, S., Ho Phuoc, T., Granjon, L., Guyader, N., Pellerin, D., Guérin-Dugué, A., 2009. Modelling spatio-temporal saliency to predict gaze direction for short videos. Internat. J. Comput. Vision 82 (3), 231–243.