Research Paper

Volume : 4 | Issue : 7 | July 2015 • ISSN No 2277 - 8179

An Efficient Segmentation and Action Recognition in Video Using Visual Saliency and Reliable Region Approach

Computer Science

KEYWORDS : Visual Saliency, Reliable Region, Action Recognition, k-means

Dr.L.Sankari — Associate Professor, Department of Computer Science, Sri Ramakrishna College of Arts and Science for women, Coimbatore, Bharathiar University, India

S.Rekha — Research Scholar, Department of Computer Science, Sri Ramakrishna College of Arts and Science for women, Coimbatore, Bharathiar University, India,

ABSTRACT

A novel method for video segmentation is proposed for detection of small sized classes and to reduce the computational burden of algorithms. Maximum symmetric model approach and Exponential ellipse model are integrated to obtain the saliency value for each pixel. Secondly, the Gaussian Mixture Model (GMM) is applied to locate the object region. Visual saliency method and GMM method successfully segments the object. Then computational burden of the work can be reduced by exploring the idea of finding and using only the regions those are reliable for detection. To reduce the computational burden, object class segmentation is done with reliable regions. In addition, actions are recognized for the input images. First the color moment features are extracted from both training set and test data. Then the obtained features are compared with the training set.

INTRODUCTION

Video segmentation has a number of interesting applications, including video editing; harvesting labelled video data for training classifiers and learning shape, actions as well as developing priors for unsupervised video segmentation. In the past, several heuristic systems for video segmentation have been proposed which process a few frames at each step. In recent times, video segmentation has gained a lot of attention especially as extensions of image super-pixellization to space-time super-pixels. The aim of these methods is to group pixels which are photo metrically and motion wise consistent. In simple cases, where there is a clear distinction between foreground and the background, the grouping may appear to be semantically meaningful. However, in more complex videos, the result in general is an over-segmentation, and requires additional knowledge to achieve any object level segmentation. The issues regarding surrounding small sized classes and ease the computational burden of the algorithm are not considered. The actions are also not recognized in this work.

PROPOSED WORK:

The following are the proposed approach used to segment an object from the background. This system is used to reduce the complexity and the actions are recognized. The first proposed method called visual saliency is used to segment the region and the other method is used to reduce the computational burden.

PROPOSED METHOD 1: OBJECT SEGMENTATION USING VISUAL SALIENCY

Visual saliency is a property which makes an object, a person, and pixel stand out in relation to its neighbours and thus, capture the viewers attention. Two approaches are considered here for visual saliency detection, they are maximum symmetric surround model and exponential ellipse model. When certain filters are passed on the image to obtain the saliency, the problems which may occur are highlighting the salient object may fail and undesired salient regions are highlighted when the background is complex.

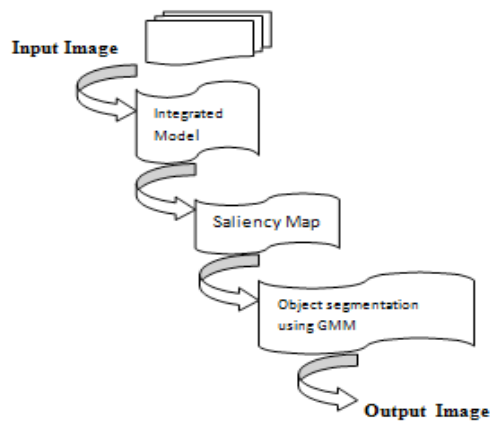

Figure 1: Visual Saliency Map Construction

- **Maximum Symmetric Model**

symmetric model approach can solve these problems by varying the bandwidth of centre-surround filter. This means that the bandwidth of this filter is very large in the center of images and becomes smaller at the borders. The limitation still exists in this method, when it fails to evaluate the saliency value for each pixel by considering the distance between each one of the borders of the image precisely. For example, if the image borders cut the salient object, they are treated as the background and are less likely to be detected.

- **Exponential Ellipse Model**

Exponential ellipse model considers the high saliency value for the pixel in the center of image and decreases the amount of pixels that surround the borders by an exponential level. The above two methods are integrated to obtain the saliency value for each pixel, which conveys the information about the distance from each pixel to each border and to the center of the image. The saliency value for each pixel by a model is used that combines a maximum symmetric surround model with an exponential ellipse saliency model. In general, the formula for calculating the saliency value [1] at the given pixel is as follows:

$$s\left(1 - e^{\left(\frac{x-\rho(i)}{R}\right)^2 + \left(\frac{y-\rho(j)}{c}\right)^2}\right) \| I_{\mu}[x,y] - I_{f}[x,y] \| \quad (1)$$

where,

- **x,y** are pixels
- **ρ (i) and ρ (j)** is the location of the pixel in the center of image that follows the x-axis and y-axis, respectively.

- **(R, C)** is the size of the input image.
- $\mathbf{I_\mu(x, y)}$ is the average of the sub image at the center pixel(x, y)
- $\mathbf{I_f}$ is the blurred image.

- **Segmentation for objects of attention based on the GMM**

The Gaussian Mixture Model (GMM) is popular for color clustering and image segmentation. The most important issue for GMM implementation is to get suitable parameters. For an input image, first the visual attention saliency map is calculated to locate the rough region of the objects of attention. Then the GMM is used to segment the objects from the rough attention region.

PROPOSED METHOD 2: RELIABLE REGION

For robustness, a decision has to be made on the reliability of a whole group of similar segments, rather than decide for each segment separately. An image is segmented into multiple segments. A single object class is usually segmented into several possibly overlapping parts, sometimes bearing little similarity to each other. Segment clustering is performed within each object class.

- **K Means Algorithm**

Clustering of similar segments is obtained by k-means algorithm. The reliable clusters are obtained (easily recognizable parts of the object class). The given data set are classified into certain number of clusters. The main idea is to define k centroids, one for each cluster. These centroids are placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point k new centroids are re-calculated as barycenters of the clusters resulting from the previous step. After these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop the k centroids change their location step by step until no more changes are done.

The reliable clusters for each object class are obtained in the training class. The distance is computed from the current segment to the cluster centers obtained at the training stage. If the current segment is closest to an unreliable cluster center, it is itself deemed as unreliable and discarded. Otherwise it is classified using the k-nearest neighbor classifier.

Since unreliable segments are discarded for a test image, there is a chance that a pixel does not get classified because it is contained only in unreliable segments. Experimentally the percentage of such pixels is small. In most cases, if a pixel is in unreliable generic blob in one segmentation, it often does belong to a reliable region for segmentation, usually at a different scale. The small percentages of pixels that do not get classified at this stage do get classified at the next stage. Since all the segments are classified independently, to improve coherence and integrate the results of multiple segmentations, graph cut optimization framework is used. Generally a final labelling (f) is given as [2],

$$\mathbf{E(f)=\sum_{p\in P}\ D_p\ (f_p)} \qquad (2)$$

The first term in equation is called a data term and it is the penalty for pixel p to be assigned label f_p. D_p measures how well label f_p suits pixel p.

ACTION RECOGNITION

In addition actions are also recognized in the proposed work. First the features are extracted from training set and test data. Then the obtained features are compared with the training set.

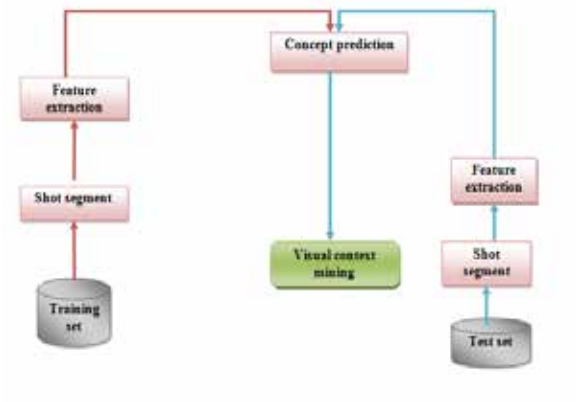

Figure 2: Action Recognition

- **Observed Feature**

In this module, features of the shots are considered. A feature is a piece of information which is relevant for solving the computational task related to a certain application. In both training and test data the feature should be extracted.

The color moments are extracted using mean and standard deviation which are mentioned below.

Mean:

First color moment can be interpreted as the average color in the image.

$$E_i = \sum_{j=1}^{N} \frac{1}{N} P_{ij} \qquad (3)$$

Where, N is the number of pixels in image.

- P_{ij} is the value of j^{th} pixel of the image at the i^{th} color channel.

Standard Deviation:

SD is the second color moment obtained by taking the square root of the variance of the color distribution.

$$\sigma_i = \sqrt{\left(\frac{1}{N}\ \sum_{j=1}^{N} (P_{ij} - E_i)^2\right)} \qquad (4)$$

- E_i is the mean value or first color moment for i^{th} color channel.

- **Concept Prediction**

Pearson product moment correlation is used to measure strength of relationship between concepts C_j and C_k , whose definition is given as follows:

$$PM(c_j, c_k) = \frac{\sum_{i=1}^{|S_{trn}-1|}(y_i^j - \mu_j)(y_i^k - \mu_k)}{|S_{trn}-1|\sigma_j\ \sigma_k} \qquad (5)$$

Where S_{trn} is the training set μ_j and σ_j are sample mean and standard deviation of observing concept C_j in training set S_{trn} respectively.

- **Visual Context Mining**

To achieve better performance, the visual context information is used to refine the target concept, which gives a second chance to optimize the predicted probabilities of the shots according to the visual similarities among them.

RESULT AND DISCUSSION

The datasets for proposed system are taken from the SegTrack datasets. Since, the images are animated the number of frames gets vary for each dataset. The actions for the dataset are also determined. The performance metric of Accuracy, precision, recall and f-measure are estimated for the below dataset.

Volume : 4 | Issue : 7 | July 2015 • ISSN No 2277 - 8179

TABLE – 1
HUMMING BIRD COMPARATIVE TABLE

	Existing system	Proposed system	
Metric	Graphical Model	Visual Saliency	Reliable Region
Accuracy	90.449	92.308	96.923
Precision	0.90149	0.92564	0.96923
Recall	0.90828	0.92443	0.96968
FMeasure	0.90483	0.92503	0.96945

The above dataset belongs to the SegTrack dataset. The performance metric for accuracy, precision, recall and F-measure are shown in the above table. Among the entire proposed methods Reliable region shows better performance.

CONCLUSIONS

In the present work, two methods for video segmentation are proposed for the detection of small sized classes and to reduce the computational burden of algorithms. The proposed method 1 visual saliency is used to detect small sized objects in frames and GMM is used to segment the object. The proposed method 2 reliable region is used to reduce the computational burden by exploring the idea of finding and using only the regions those are reliable for detection by using simple nearest neighbor classifier. This discards unreliable regions results and significantly shows improvement and thus reduces the computational burden of the segmentation process.

In addition the actions are recognized for the given video. The color moment features are extracted from RGB using mean and standard deviation. The features from the test set are compared with the training set using Pearson product moment correlation. Visual context mining is used to refine the results obtained from the previous step with the training set.

In the proposed work the actions are recognized only for the selected image. In Future, this work may be extended to all types of images. If there is no action part selected the future work may recognize as 'NO ACTION'.

REFERENCE

[1] Huynh Trung Manh and Gueesang Lee, "Small Object Segmentation Based on Visual Saliency in Natural Images", Proc. J Inf Process Syst, December 2013. | [2]B Boykov.Y, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," Proc. Seventh IEEE Int'l Conf. Computer Vision, 1999. | [3] Ayvaci.A, M. Raptis, and S. Soatto, "Sparse Occlusion Detection with Optical Flow," Int'l J. Computer Vision, vol. 97, pp. 322-338, 2011. | [4] Badrinarayanan.V, F. Galasso, and R. Cipolla, "Label Propagation in Video Sequences," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2010. [28] | [5] Bai.X, J. Wang, D. Simons, and G. Sapiro, "Video SnapCut: Robust Video Object Cutout Using Localized Classifiers," Proc. ACM Siggraph, pp. 70:1-70:11, 2009.