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

Usability of Cluster Based Co-Saliency in Video Foreground Detection

Merin Joseph[1], Anil A. R.[2]

[1]M.Tech Scholar, Department of Computer Science and Engineering, Sree Buddha College of Engineering, Alappuzha, India

[2]Associate Professor, Department of Computer Science and Engineering, Sree Buddha College of Engineering, Alappuzha, India

Abstract: ***Ability of human visual system to detect prominent regions in an image is fast, reliable and efficient. Computational modeling of this extraordinary behavior is termed as saliency detection. Saliency detection identifies salient regions in an image and is relevant in many computer vision applications such as object recognition, image segmentation, image retrieval etc. The term co-saliency can be explained as identifying common prominent regions in multiple images. In this paper we analysis various saliency and co-saliency detection mechanisms and studies the usability of cluster based co-saliency detection method in video foreground detection. The paper also explains design and implementation of cluster based co-saliency in video foreground detection.***

Keywords: Saliency, Co-saliency, Clustering, K-means, Video Foreground Detection

1. Introduction

Extracting information from still images and videos is one of the most important and complicated task in computer vision and image understanding. Identifying region of interest in an image have two approaches, i.e. top-down approach and bottom-up approach. In top-down approach we look for specific objects. Prior knowledge about the target object is known. Such an approach comes under object recognition algorithms. On the other hand in bottom up approach we do not have prior knowledge of the region to be detected. Such class of algorithm tries to detect regions which are prominent or salient i.e. the region that stays unique from other regions and simultaneously capture attention. Hence the detection in this case is generic as there is no specific target object. Such class of algorithm is called saliency detection algorithm. The paper presents various aspects of saliency detection mechanisms.

The paper discusses co-saliency a less explored area which identifies common saliency on multiple images. Visual co-saliency is a subjective perceptual ability that makes similar objects in a group of images that captures our attention by visually co-salient stimuli. Thus co-saliency region have two properties. A) A co-salient region in an image should be locally salient i.e. it should be a prominent region in the image. B) A co-salient region should exhibit high similarity between multiple images. In our paper we focus on a cluster based co-saliency detection technique and its usability in video foreground detection.

Finally the paper describes the design and implementation of cluster based co-saliency in video foreground detection. Our system works well in extracting simple single object video foreground. The implementation results are also shown.

2. Background

2.1 Saliency

Saliency is the process of extracting visually captive regions or objects in an image and thus it is closely related to human perception or how human process the visual stimuli. An example of saliency detection is as shown in fig 1 .Many studies were made in multiple disciplines to model this human perceptual ability. Computer vision had proposed many methods to model this basic intelligent behavior.

Figure 1: Saliency Detection

L. Itti, C. Koch, and E. Niebur [1] proposed a model of saliency based visual attention for rapid scene analysis. It identifies saliency of image by evaluating central surrounded differences across multi scale image features. Here various spatial scales of input image are created and by extracting multiple visual features such as color, intensity, orientation etc. to detect saliency. The method works well in identifying saliency based on implemented features say color, intensity and orientation but fails in detecting saliency caused by other image features.

Independent of features or any other prior knowledge of the image content, prominent region in image is detected in certain saliency detection methods. A spectral residual approach [2] constructs a log spectrum of the input image, analysis it and extracts the spectral residue of an images in spectral domain. It thus introduces a fast method of constructing saliency map in spatial domain based on the spectral evaluation. Frequency tuned salient region detection

[3] extract saliency from the frequency domain analysis of the input image. It identifies the relevance of spatial frequency in saliency detection and introduces a frequency tuned approach of computing saliency. The above algorithms are fast and usable but not efficient and reliable in their detection results.

It is known that high contrast visual signals stimulate human visual receptive cells and hence contrast property in images is extensively used in many saliency detection algorithms. Contrast-based image attention analysis by using fuzzy growing [4] introduces a local contrast analysis for saliency detection which is extended further using a fuzzy growth model. A Global contrast based salient region extraction [5] computes saliency value of each image pixel based on its contrast and spatial coherence with every other pixel. It uses a histogram based method to calculate the pixel saliency. The method is efficient in its detection results.

2.2 Co-saliency

While considering multiple images common prominent regions or object may occur in these images. Co-saliency deals with detecting this common saliency in multiple images as shown in fig.2. Co-saliency detection is a less explored area which is usable in many computer vision applications such as co-segmentation, images retrieval etc.

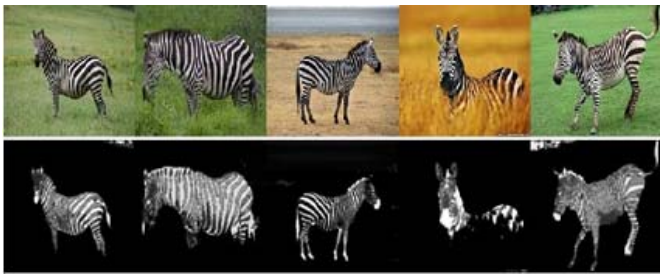

Figure 2: Co-saliency Detection

Co-saliency detection is firstly defined in [6] as identifying the common unique region in a group of images. The method was implemented based on a user study. From this user study data, a model was trained for image saliency in context of other images that we call as co-saliency. Its method of implementation which involves training is not usable in automatic detection of co-saliency.

Pre-attentive co-saliency detection [7] proposes a mechanism to find co-saliency between image pairs by enhancing similar and pre-attentive patches. A co-saliency model for image pairs [8] introduced co-saliency as a linear combination of the single image saliency map and multi image saliency map by employing a complex co-multilayer graph. Major disadvantage these models [7] [8]is that they were designed for image pairs and is hard to generalize to the case of multiple images.

Cluster-based Co-saliency detection [9] introduces a new cluster based detection mechanism which made use of contrast property and spatial property to detect uniqueness in images and a correspondence evaluation between multiple images to detect similarity between them. In this method the co-saliency map for each image is implicitly identified through the clustering process. Here a two level clustering i.e. an intra-image clustering and an inter-image clustering is carried out to classify the pixels of each image in to clusters based on their gray level values. Hence it uses three detection cues for defining saliency for each cluster, a contrast cue, a spatial cue and a correspondence cue.

2.3 Video Foreground Detection

Video foreground detection approaches can be of supervised or unsupervised. Most of the existing methods relay on supervised pre-learned model or unsupervised background model which makes the implementation complex and imperfect. For example [10] decomposes object shape in hierarchical way and train the system to detect all possible configurations of the target region of interest. Another example [11] uses an interactive scheme which requires users to provide manually the ground truth label information.

Most of the unsupervised methods uses background model [12][13]which has many difficulties such as (1) slowly-moving foreground regions may incorrectly be labeled as background, (2) occluded background may be wrongly classified as foreground, (3) changing illuminations, which are common in application scenarios, often corrupt the motion information. Some unsupervised algorithm uses features associated with foreground objects for foreground extraction. It uses complex graph based methods [14]. Another approach made use of saliency in sequence of frame [15] by calculating visual saliency and motion saliency by considering the optic flow map.

3. Video Foreground Detection using Cluster Based Co-saliency

The Cluster based co-saliency detection is a bottom–up method of detecting common saliency in multiple images without any heavy learning. Being simple and efficient it can be easily used as a pre-processing method in many applications such as co-segmentation, image retrieval etc. Beyond this another application of co-saliency is simple single object video foreground detection.

Here we suggest a simple method of detecting video foreground using cluster based co-saliency detection method. We know that a foreground of an image can be defined as the noticeable region or salient region in that image. Also in simple natural videos a foreground object will repeat in sequence of frames. Thus co-saliency can be effectively used for detecting this foreground since it can extract common saliency. Thus our method proposes simplest way of detecting video foreground. Our method eliminates complex graph calculations and motion analysis of subsequent image frames.

4. Design and Implementation

The work flow design of our video foreground detection system is as shown in Fig.3. As shown the input video is first converted in to sequence of image frames. Then clustering of

images is performed and subsequent calculation of various saliency values for clusters is done. Finally obtained co-saliency map is used of reconstructing the video with specified foreground. Thus our video foreground detection system design involves of the following modules.

- Clustering module
- Contrast cue module
- Spatial cue module
- Corresponding cue module
- Mapping module

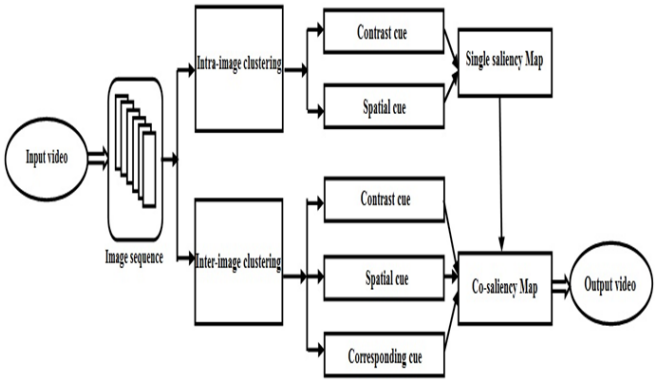

Figure 3: System Work Flow

4.1 Clustering module

Given a video file, after converting to sequence of frames we employ a two level clustering of image sequence based on its gray level values. It includes an Intra-image clustering and an Inter image clustering. Intra image clustering clusters each image separately and inter image clustering is done by associating the pixels on all images and classifies into clusters. Inter image clustering i.e. clustering on multiple images provides global corresponding relationship between images.

The *k*-means clustering method can be used to classify image pixels. Our system used a *k*-means ++ algorithm which is an enhancement to *k*-means clustering. The *k*-means++ algorithm [16] proposes a procedure to initialize the cluster centers before proceeding with the standard *k*-means optimization iterations. With the *k*-means++ initialization, it is guaranteed to find a solution that is O(log *k*) competitive to the optimal *k*-means solution.

Here a sequence of M images is considered and each pixel P_{ij} is classified in to k clusters. For intra image clustering we chose cluster number as six and in inter image clustering the cluster number is chosen as min $\{3M, 30\}$. The cluster C_k represents the k th cluster and a saliency value is calculated for each cluster.

4.2 Contrast cue module

Contrast is a property that stimulates the human visual receptive fields and thus contrast cue can effectively used for extracting visual uniqueness in images. Also calculating contrast cue in inter image clusters contribute to correspondence between multiple images.

The contrast cue $w^c(k)$ of cluster C^k is defined using its feature contrast to all other clusters.

$$w^c(k) = \sum_{i=1,i\neq k}^{K} \left(\frac{n^i}{N} \| \mu^k - \mu^i \|\right) \quad \text{--- (1)}$$

where μ is the cluster center contrast n^i is the total number of pixels in cluster C^i and N is the pixel number of all images.

This results in larger clusters with high contrast to play more influence.

4.3 Spatial Cue Module

Human visual system pays more attention to image center than other region. This is explained by center bias rule.ie, when the distance between object and the image center increases, the attention gain is depreciating. To measure this spatial property at cluster level we use the following equation.

The special cue $w^s(k)$ of cluster C^k of M images is defined as:

$$w^s(k) = \frac{1}{n^k} \sum_{j=1}^{M} \sum_{i=1}^{Nj} \left[N\left(\left\| z_i^j - o^j \right\|^2 \middle| 0, \sigma^2 \right) . \delta\left[b(p_i^j) - C^k \right] \right]$$

---- (2)

Where $\delta(.)$ is the Kronecker delta function and Gaussian kernel N (.) computes the Euclidean distance between pixel z_i^j and the image center o^j , the variance σ^2 is the normalized radius of images. And the normalization coefficient n^k is the pixel number of cluster C^k .

Like contrast cue, spatial saliency is calculated for both intra image and intra image clusters. We know contrast cue identifies the most salient regions in an image, but spatial cue eliminates the salient or textured background especially those away from image center.

4.4 Corresponding cue module

Corresponding cue is the most important measure of calculating common saliency on multiple images. It stands for the repetitiveness i.e., how frequently the object reoccur in multiple images. It calculates the corresponding saliency for each cluster by measuring how the clusters distribute on multiple images. Thus it uses a M bin histogram for each cluster.

The M bin histogram $\hat{q}^k = \{\hat{q}_j^k\}_{j=1}^{M}$ describes the distribution of cluster C^k in M images:

$$\hat{q}_j^k = \frac{i}{n^k} \sum_{i=1}^{N_j} \delta\left[b(p_i^j) - C^k \right] \text{ , j=1....M} \quad \text{---- (3)}$$

Where n^k is the pixel number of cluster C^k which enforces the condition $\sum_{j=1}^{M} \hat{q}_k^j = 1$.

Then corresponding cue is defined as:

$$w^d(k) = \frac{1}{\text{var}(\hat{q}^k) + 1} \quad \text{---- (4)}$$

Where $\text{var}(\hat{q}^k)$ denotes the variance of histogram $\hat{q}^k$ of the cluster C^k

4.5 Mapping module

Contrast, spatial, and corresponding saliency is thus calculated and now we need to combine these saliency values. For this multiplication is employed and the cluster level co-saliency of a cluster C^k is defines as,

$$p(C^k) = \prod_i w^i(k) \quad \text{---- (5)}$$

Now we have cluster level co-saliency values which provide a discrete assignment. We smooth the co-saliency values for each pixel using the Gaussian distribution N .Now we detect the co-saliency in images sequence. To obtain the result as video foreground, the resulting image frames are again converted to video file.

5. Result

The system was successfully implemented and observed that it works efficiently in detecting video foreground of simple single object videos. The implementation result on a video sequence is shown in Fig. 4.

Figure 4: Video Foreground Detection

6. Conclusion and Future scope

We presented a video foreground detection using cluster based co-saliency. A simple video always contains common saliency in its frames which is its foreground. Thus co-saliency detection can effectively used in foreground detection. A cluster based method of approach makes it easy to implement and usable. It uses contrast and spatial property simultaneously to detect uniqueness in a frame. The implementation result shows the effectiveness and reliability of our method. In future we planned to include texture properties to improve detection results of complex videos.

References

[1] L. Itti, C. Koch, and E. Niebur,"A model of saliency-based visual attention for rapid scene analysis", IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 11, pp. 1254-1259, 1998.

[2] X. Hou and L. Zhang,"Saliency detection: a spectral residual approach," IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[3] R. Achanta, S. S. Hemami, F. J. Estrada, and S. Ssstrunk,"Frequency tuned salient region detection," IEEE Comp. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), 2009,

[4] Y.-F. Ma and H.-J. Zhang. "Contrast-based image attention analysis by using fuzzy growing."ACM Multimedia,

[5] M. Cheng, G. Zhang, N. J. Mitra, X. Huang, and S. Hu, "Global contrast based salient region detection," in Proc. Comput. Vis. Pattern Recognit., 2011,

[6] D. Jacobs, D. Goldman, and E. Shechtman," Cosaliency: where people look when comparing images," in ACM symposium on User interface software and technology, 2010, pp. 219-228.

[7] H. Chen, "Preattentive co-saliency detection," in ICCP,2010,pp.1117-1120

[8] H. Li and K. Ngan,"A co-saliency model of image pairs," IEEE Trans. Image Process., vol. 20, no. 12, pp.

[9] Huazhu Fu, Xiaochun, Zhuowen Tu,"Cluster-based Co-saliency Detection" IEEE Trans. Image Processing, vol 30, pp 330-342, 2013.

[10] Z. Lin and L. S. Davis, Shape-based human detection and segmentation via hierarchical part-template matching, IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 4, pp. 604618, Apr. 2010.

[11] Y. Y. Boykov, O. Veksler, and R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 11, pp.

[12] T. Bouwmans, F. E. Baf, and B. Vachon, Background modeling using mixture of Gaussians for foreground detection: A survey, Recent Patents Comput. Sci., vol. 3, no. 3, pp. 219237, 2008.

[13] F.-C. Cheng, S.-C. Huang and S.-J. Ruan, Advanced background subtraction approach using Laplacian distribution model, in Proc. IEEE Int. Conf. Multimedia Expo, Jul. 2010, pp. 754759.

[14] K.-C. Lien and Y.-C. F. Wang, Automatic object extraction in single- concept videos, in Proc. IEEE Int. Conf. Multimedia Expo, Jul. 2011, pp. 16.

[15] Wei-Te Li, Haw-Shiuan Chang, Kuo-Chin Lien, Hui-Tang Chang, and Yu-Chiang Frank Wang, "Exploring Visual and Motion Saliency for Automatic Video Object Extraction", IEEE trans. on Image Processing, vol.22,No.7, Jul 2013

[16] David Arthur and Sergei Vassilvitsk "k-means++: The Advantages of Careful Seeding", Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms.