Video Object Detection based on Region Contrast

Matheel Emaduldeen Abdulmunim, Ph.D. (Assist. Prof.)*

Abstract: Foreground object detection is one of the major impotent tasks in the field of computer vision which attempt to discover important objects in still image or image sequences or locate related targets from the scene. Foreground objects detection is very important for several approaches like object recognition, surveillance, image annotation, and image retrieval, etc. In this work, a new method was proposed for detection and segmentation foreground object from image or video in case when the targets are moving or not moving. The proposed method are able to detect important target for case the target is moving or not and can separate foreground object with high details. Comparisons with general foreground detectors such as background subtraction techniques were done and conclude that the proposed method has a high accuracy with detection rate 98.05.

Keywords: Image Segmentation, Foreground Object, Background Subtraction.

* University of technology, Baghdad, Iraq.

1. Introduction

Biological vision systems tend to be incredibly efficient at finding relevant targets within a scene. Most people will probably quickly and consistently spot at those important targets within the images for example in Figure 1. Certainly, to detect these types of objects from image, it commonly preselected by using a two level labeling strategy to be sure a foreground object is detached from the sense background [1].

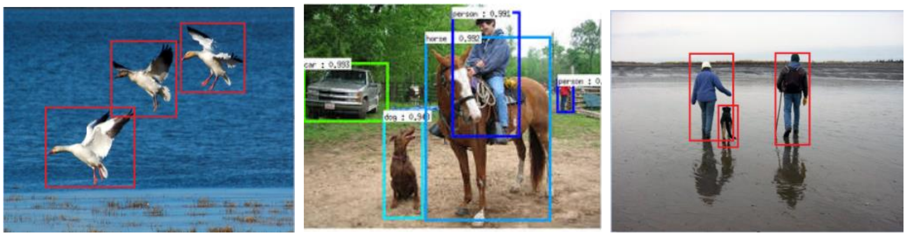

Figure 1. Sample of images having an obvious forgroubd object, where the rectangles represent an object annotation.

Recognizing these areas of prominent, or significant, within the visual field allows one to recruitment the minimal perceptual resources using an efficient way. When compared computer vision approaches with biological systems, it's certainly the computer vision techniques are far behind capability to detect foreground objects. Nevertheless, dependable foreground object detection strategies could be beneficial in numerous applications such as: object recognition, unsupervised image segmentation, and adaptive scaling and compression. One common strategy to minimize scene clutter would be to recognize foreground objects versus a static background. There are many techniques currently successful in several purposes that useful in detect moving object in the scenes such as background subtraction, but the difficult task is to detect the foreground object from moving sense or still image that can focused and segment just those important objects from other objects in background of sense [2, 3].

The automated recognition of foreground object regions in images requires a gentle breaking down the foreground from other elements of background image. This type of breaking down is a key element of several graphics and computer vision tasks. Instead of emphasizing forecasting human fixation points (an alternative significant research path of visual attention modeling), foreground region detection strategies focus on regularly featuring existing foreground object regions, and as a result benefiting numerous applications, such as object recognition, object level image manipulation, object of-interest, internet visual media retrieval, image segmentation, content aware image editing and adaptive compression [4].

Extraction of salient or foreground objects in a scene relates to appropriate object retrieval and image segmentation. Apparently, reliable foreground evaluation is frequently achievable without the need of actual scene knowing. Foreground, considered a bottom-up procedure that derives from visual surprise, rarity or distinctness, and is often related to variants in image characteristics like gradient, color, boundaries and edges. Visual foreground objects are investigated throughout numerous disciplines such as computer vision, cognitive psychology, and neurobiology. According to observed reaction along biological methods, the human attention theories hypothesize that the techniques of human vision system basically areas of an image in detail, along with leaving the remaining around unprocessed. Earlier work suggests two steps of visual consideration: a quick, data driven, bottom-up, pre-attentive, foreground object extraction; then, slow, top-down, goal driven, task dependent, foreground extraction [5].

2. Previous Work

A different method has been provided for foreground and silence object detection approaches, which have been aimed to locating specific categories such as tables, cars, persons, airplane, etc.

Walther and Koch (2006) [6], specify proto objects as peaks spatial format of the foreground map. Basically, their method suggests that the proto object in image has a set of pixels that is determined by a continuing four linked neighborhood of the peak with foreground over a specific threshold. Hence, in their method, the majority of salient points are determined based on the spatial-based model, following that the foreground is distributing on the area around them, which mean that there are proto objects has been obtained from the foreground map.

Liu et al. (2007) [7], consider color spatial distribution, center-surround color histograms, and multi-scale contrast to evaluate pixel foreground. For the determination step, overall characteristics are summed in a depending different field causing a binary labeling strategy that isolates the foreground object away from its background. This technique proves good results. Nevertheless, it contains the Conditional Random Fields (CRF) that is generally computationally high priced.

Valenti et al. (2009) [8], presented a salient object detector in real time. With their approach, pixel foreground has been computed as a matrix of different characteristics: curvedness, rarity and is centricity of color lines. This method demonstrates edges and centers of the image constructions. To be able to distribute foreground objects inside connected regions, the researcher attempts average values of the foreground map and graph based segmentation inside each segment. In a localization step, effective sub-window search has been used.

Bruce and Tsotsos (2009) [9], outline that foreground depending on optimum information simples. They calculated the Shannon self-information via using the possibilities of the image component within a patch provided the component of the entire image. Patches having unusual component tend to be very informative, and therefore salient.

Marchesotti et al. (2009) [10], suggest a detector for salient object that is depending on the hypothesis the images with a same visual scene will probably have salient objects with similar characteristics. In order to determine foreground inside a final image, the researchers train a classifier for the almost all identical images, with presented ground real boxes close to foreground objects. A two type problem for classification has been considered: non-salient class includes the background and the salient class contains salient objects. Every patch in the target image is categorized for being non-salient/salient. To be able to discover a salient object, the classifier output has been used to initialize the algorithm of iterative graph-cut. Therefore, the segment which provides coverage for many of the salient pixels is described to be salient object. The approach has been proven to acquire highly good results whenever annotated image data is obtainable. Nevertheless, the researchers additionally demonstrated how the technique is greatly dependent upon the grade of the retrieval step where the almost all related images are extracted.

Van de Sande et al. (2011) [11], has been modify hierarchical segmentation to locate beneficial candidates of object locations. Their work is dependent on a couple key suggestions. Firstly, objects could be of any size that can appear at any scale. As a result, a hierarchical segmentation method has been used and all segments all over the entire structure are considered. Secondly, to be able to take into account various object performances and image conditions, the outcome of numerous, complementary segmentations is integrated. This approach has shown itself powerful task for object localization.

3. Foreground Detection Theory

A different method has been provided for foreground and silence object detection approaches, which have been aimed to locating specific categories such as tables, cars, persons, airplane, etc. [12].

There are two popular hypotheses for human detection: object-based and spatial-based detection. For the spatial-based hypothesis, interest is over a zoom lens or a spotlight that shifting our focus from one spatial spot to one other to sample surrounding. Consequently, all visual content inside a fovea-sized region close to those locations has been processed.

The idea of many foreground and silence object detection algorithms are able to go back towards the feature detection strategy that detect that varieties of interest are dependable with regard to joining different features into knowingly experienced wholes. So, a model of detection method has been built based on a biologically credible architecture. It symbolizes the input image via the color, orientation, and intensity parts, as well as can determine foreground maps by making use of local around differences that are used to make a final foreground map. In recent times, many researches have been developed to design numerous foreground features characterizing foreground objects or regions [13, 14].

The majority of research generally follow the contrast (or center-surround difference) framework. In center-surround theory, the color histograms, calculated to present the center and the surround, which are utilized to find center surround dissimilarity. The information hypothesis standpoint is subjected to provide a mathematical formula, calculating the center surround divergence depending on feature information. The framework “center surround difference” could be examined to determine the foreground from region-based image information. The variance that is between a straightaway neighboring regions and color histogram region are utilized to analyze the foreground rating. The global contrast approaches, computing the foreground map by comparing all area with each other's, intend to immediately calculate the global uniqueness. Using the regional contrast, feature color uniqueness and spatial distribution are brought to calculate the foreground ratings of regions. The foreground map is made by propagating the regions of foreground rates to the pixels [15].

The latest research for detect foreground/silence objects are using statistically modeling, isometric color and curvedness, making use of multitask sparsely pursuit, depth cues, implementing image histogram, combing top-down and bottom-up features, quaternion-based spectral analysis, task-specific visual attention, exploring patch rarities, and so on. There are several other foreground object detections, such as context-aware foreground detection that attempting to discover the image regions that present the scene [16].

4. Proposed Method

This paper focused to detect the important objects and separate it from moving and still sense with high details in order to increase the accuracy of recognition process. This paper proposing an approach for detects and segment foreground object depending on region contrast. Since the theory of object based attention is recommending to start using the complex image segmentation towards proto-objects. Even though, common object segmentation is a difficult task, estimated at describing significant segments within an image via feature grouping is probably achievable. Then determining foreground for each proto-object coupled with describes the foremost salient one.

This method of foreground object detection consists of four steps: abstractions of images form video, tree structure, matrix breaking down and assignment foreground object. Figure (2) gives the block diagram of this method. Object detection block can be constructed from the steps in figure (3). Algorithm (1) gives the essential steps of this method.

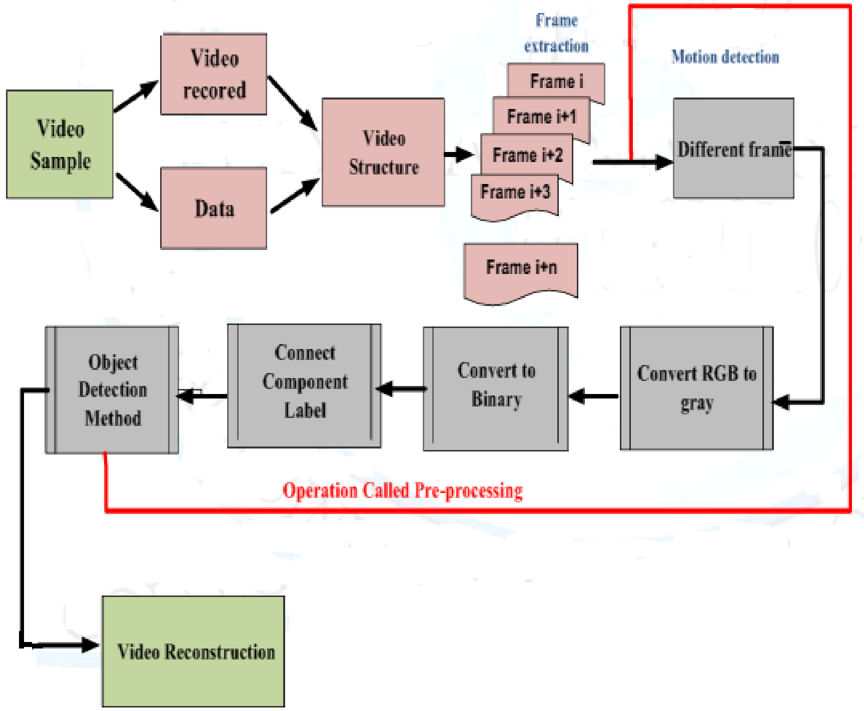

Figure 2. The whole block diagram of the method.

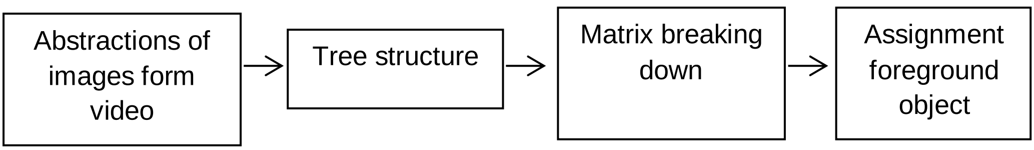

Figure 3. Object detection steps.

In step (2.3), to find the Edge Density Let i(x, y) be a window and e(x, y) be the edge magnitude of the window. For a sub region r with the left-top corner at (x1,y1) and the right-bottom corner at (x2,y2), the edge density feature is defined as:

$$f = \frac{1}{a_r} \sum_{x=x_1}^{x_2} \sum_{y=y_1}^{y_2} e(x,y) \qquad (1)$$

where a_r is the region area, a_r = (x2- x1+1)(y2- y1+1). Algorithm (1) shows the systematic steps of the work.

Algorithm 1: **Foreground objects detection.**

Input: Color video.

Output: Assignment foreground objects for color image (I).

Begin

Step1: Separate the input video into frames. For each frame (image (I)) do the following.

Step2: Preprocessing:

Step 2.1: Convert image (I) to grayscale image (J).

Step 2.2: Remove image frame (edges) for top, bottom, left and right of image (J).

Step2.3: Set the maximum width of frame then evaluate the Edge Density (ED) by equation (1).

Step3: Image Segmentation:

Step3.1: Use Simple Linear Iterative Clustering (SLIC) algorithm (segment it to different patches).

Step3.2: Convert each batch as a feature vector FV [].

Step3.2: Constitute feature matrix FM [] from FV []s.

Step4: Extract features:

Step4.1: Conclude small and perceptually homogeneous features (low-level features).

Step4.2: Steerable pyramids and RGB color to generate a dimension feature description.

Step5: Create index tree:

Step5.1: The super pixels are added to the index tree to encode construction information through hierarchical segmentation.

Step5.2: Calculate the appreciation of each surrounding patch by get the first and second order reachable matrix.

Step5.3: Use the algorithm “graph based image segmentation” (GBIS) to combine spatial neighboring patches based on their affinity.

Step5.4: GBIS algorithm generates a granularity-increases segmentations sequence.

Step5.5: In every granularity layer, segments are corresponding to nodes in the equivalent layer within the index tree.

Step6: Structured matrix decomposition:

Step6.1: Use the “Structured Low-Rank Matrix Factorization” to decompose matrix factorization to structured-sparse component and a low-rank component.

Step7: Post-processing:

Step7.1: Following decomposing, proceeding the outcome from the feature to some preprocessing algorithms in order to get improvements for foreground object.

Step7.2: Depending on the SM, specifying the function of straightforward foreground estimation for each patch.

Step8: After combining all patches together and executing context based propagation, the final forground map of the input image is obtained and then reconstruct the video.

End

In step (3.1) that is the first step in segmentation Simple Linear Iterative Clustering (SLIC) algorithm is an over-segmentation method, which selects cluster centers C_i from the sampled regular grid spaced S to efficiently generate super pixels. S can be described as the following Eq. (2):

$$S = \sqrt{N / K} \quad \textbf{(2)}$$

Where N is the number of the image pixels and K is the desired and settled number of the super pixels. Meanwhile, SLIC algorithm only searches for similar pixels from each cluster center to pixels for clustering within a 2S*2S setting region instead of in the entire image, so it can generate super pixels faster.

In step (5.1), the super pixels are added to the index tree to encode construction information; an illustration of this tree is given in Figure (4). In this example, the value of M equals 2, so each node contains either 1 or 2 values. The root contains values 16 and 25. Its left child points to values that are less than 16, and its right child points to values greater than 25. Finally, its middle child points to values between 16 and 25.

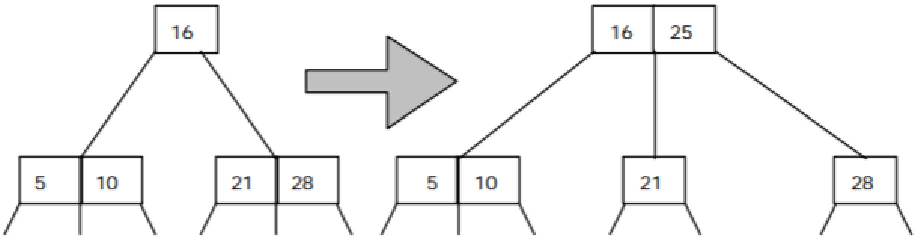

Figure 4. Index tree.

The mask shown in figure (5) (a) of size 3X3 can be used to detect the isolated pixels to construct the reachable matrix for the image that determine the connectivity between pixels to isolate the background.

The image segmentation depending on graph is instituted on choosing edges from a graph, where each pixel coincides to a node in the graph. Weights on each edge gauge the difference between pixels. The segmentation algorithm defines the boundaries between regions by contrasting two amounts – density variations toward the frontier and density variations between adjacent pixels within each region. This is beneficial knowing that the density variations across the boundary are substantial if they are large proportional to the density variations inside at least one of the regions. This effect in a method that comply confirmed non-obvious global properties. Let the internal difference of a component C in an image be

$$\mathrm{Int}\,(C) = \max w(e) \qquad \textbf{(3)}$$

Where w (e) in the largest weight in the Minimum Spanning Tree of the component.

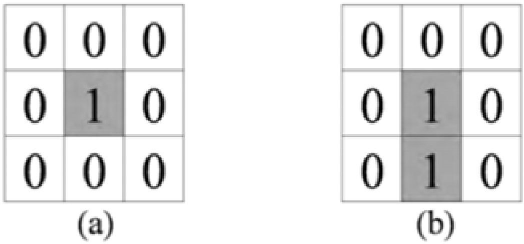

Figure 5. Masks;(a):to detect the isolated pixels, the rotation of any pixel (contour).

This model can automatically detect the distribution of foreground objects that can be found inside the connected regions. Compared, our method doesn't depend on any learning; therefore, it doesn't need image retrieval and annotation. On the other hand, it wholly depends on the object-based consideration hypothesis and extracts foreground objects straight from the image via feature grouping. That permits us to evaluate foreground at the level of proto object. Towards the best of our knowledge, it has been first apply it to task of foreground object detection. To determine integrated foreground, one has been determine the object details rarity. Automatically, the image locations that deviate via the remainder of an image needs to be foreground. To reflect the content of image, it has been employed visual color and word histograms. It has been demonstrated this foreground depending on these features outperforms standard information maximization foreground and regular spectral recurring foreground for that task of detection foreground object.

5. Results

This part shows a number of chosen results verifying the usage of the foreground object detection approach; it is useful to analyze how this approach could be successfully extract important object from background. It has been select different cases such as human, cars, jet fighters, etc. Figure (6) gives a visual image results and figure (7) gives a visual video frames results.

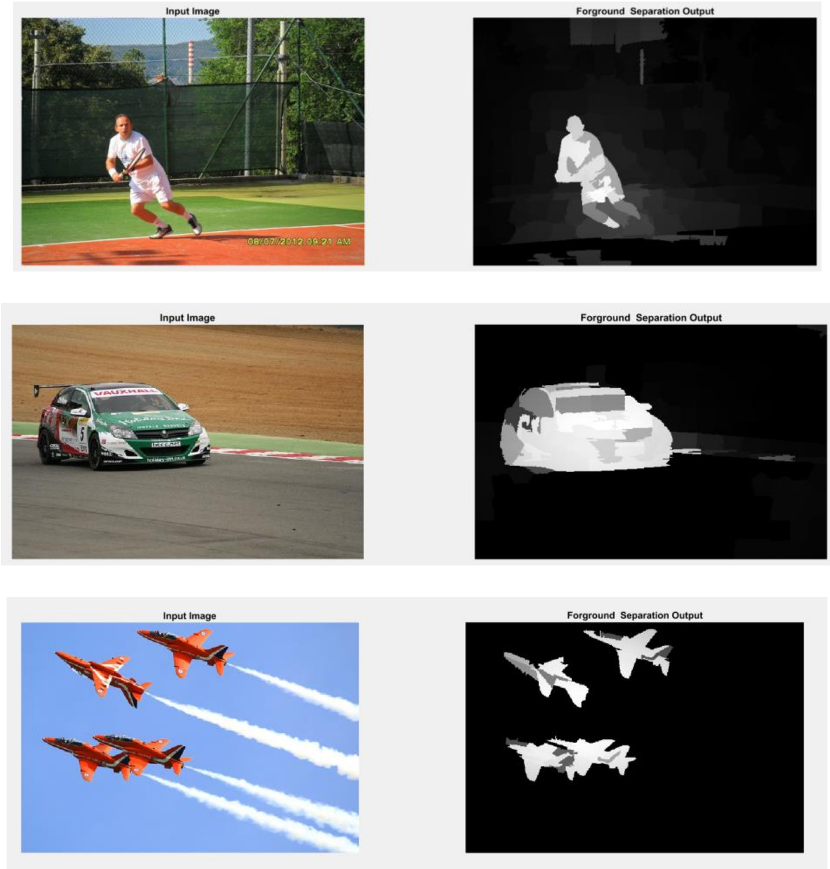

Figure 6. Prossessing Results. Left: origional image , Right: forground object detect.

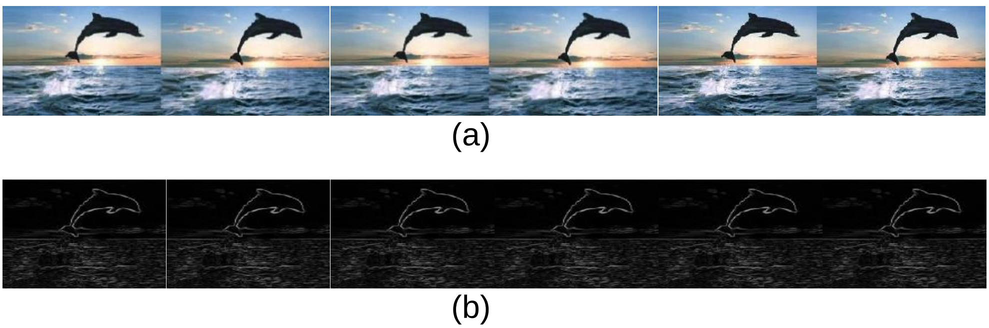

Figure 7. Prossessing Results. a: origional frames , b: forground object detection.

As shown from figure (6), it's clear that this method efficiently removes foreground object from background with high details that can be detect easily and accurately with many recognition techniques. Table (1) shows the detection rate for different videos that can be computed by dividing the no. of true detection over false detection. The average detection rate is (98.05) for different 50 videos and the average false detection are (7).

Table (1): Samples of detection rates on different videos.

Video	Number of Frames	Detection rate	Accuracy %	False detection
Doc1	1023	90	92	6
Doc2	6189	94.6	89	
Doc3	13517	96.8	94	
Doc4	20100	99.1	96	
Eng1	14578	98.9	88	8
Eng2	12453	98.5	97	
Eng3	12026	97.3	96	
Eng4	14505	95.7	94	
View1	32264	99	94	7
View2	23299	98.1	90	
View3	8775	97.7	91	
View4	7670	96.8	87	

The detection rate was compared with the Spatial-Based Model (SBM) [6] as shown in figure (8).

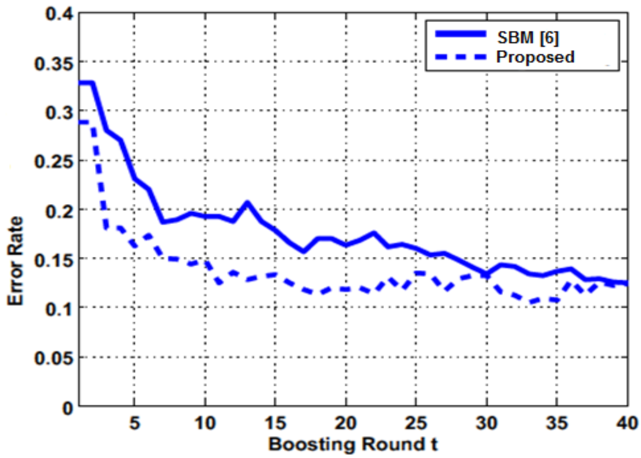

Figure 8. Error rates of different models.

6. Conclusion

In this work it has been proposed a model that dependent on the theory of object detection based attention and computes foreground of segments for image. The test provide that this method is able to separate target foreground object from sense with high accuracy and good details that gives as a numeric values in table one and figure (8) that can be recognized effectively by using recognition techniques.

The proposed method can process the pre-processing step to moving object detection, bounding object using the proposed method automatically that will implemented successfully in all frames of videos and when using a variable size of video with different resolution in addition to the different number of frames for each video, it is obvious that graph based image segmentation gives the highest rate of segmentation which proves that its work best in detecting object.The accuracy for those videos gives an average around (92.3).

Also, this method presented an object detection method based on a new image feature called edge density, which measures the presence or absence of image edges in a specific region of the object. The edge density feature can be computed very efficiently and it is found to have a better discriminative capability compared to other features, when applied to the problem of detecting people in images.

Reference

[1] Daniilidis, K.; Maragos, P.; Paragios, N.; “Computer Vision -- ECCV 2010,” 11th European Conference on Computer Vision, Heraklion, Crete, Greece, Proceedings, Part V. , September 5-11, pp.282–295, 2010.

[2] Rahtu, E.; Kannala, J.; Salo, M.; Heikkila, J.; " Segmenting Salient Objects from Images and Videos," Power Electronics, IEEE Transactions on , vol.27, no.12, pp.4858,4867, Dec. 2012.

[3] Marchesotti, L., Cifarelli; C., Csurka, G.; "A framework for visual saliency detection with applications to image thumbnailing,". In: ICCV, pp.2232-2239, 2009.

[4] Zhaoyu, P.; Pingping, L.; Changjiu, L., "Automatic Detection of Salient Object based on Multi-features," Second International Symposium on Intelligent Information Technology Application. IEEE, ISBN: 978-0-7695-3497-8, pp.437-441, January 2009.

[5] Cheng, M.; Mitra, N. J.; Huang, X.; Torr, P.; Hu, S.; " Global Contrast Based Salient Region Detection,". IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 37, Issue: 3, ISSN:0162-8828, pp.1-14, 2015.

[6] Walther, D.; and Koch, C.; "Modeling attention to salient proto-objects". Neural Networks, 19(9), pp.1395–1407, 2006.

[7] Liu, T.; Sun, J.; Zheng, N.N.; Tang, X.; Shum, H.Y.; “Learning to detect a salient object. In: IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol.33, No.2, pp.353-367, February 2011.

[8] Valenti, R.; Sebe, N.; Gevers, T.; “Image saliency by isocentric curvedness and color,”. International Conference on Computer Vision, pp.2185-2192, 2009.

[9] Bruce, N.; Tsotsos, J.; “Saliency, attention, and visual search: An information theoretic approach,”. Journal of Vision, 9(3):5, pp.1-24, 2009.

[10] Marchesotti, L., Cifarelli, C.; Csurka, G.; “A framework for visual saliency detection with applications to image thumbnailing,” In: International Conference on Computer Vision, pp.2232-2239, 2009.

[11] Van de Sande, K.; Uijlings, J.R.; Gevers, T.; Smeulders, A.; “Segmentation as Selective Search for Object Recognition,”. International Conference on Computer Vision, pp.1879-1886, 2011.

[12] Yanulevskaya, V.; Uijlings, J.; Geusebroek, J.M.; “Salient object detection: from pixels to segments”, Preprint submitted to Image and Vision Computing, pp.1-28, October 2012.

[13] Achanta, R.; Shaji A., Smith K.; Lucchi, A.; Fua P., Susstrunk, S.; “SLIC Superpixels”, EPFL Technical, Report 149300, pp.1-15, June 2010.

[14] Zujovic, J.; “Perceptual Texture Similarity Metrics”, Phd Thesis, Northwestern University, August 2011.

[15] Felzenszwalb, P.; Huttenlocher, D.; “Efficient graph-based image segmentation,” IJCV, Volum 59, No.2, pp.167-181, 2004.

[16] Haeffele, B.D.; Young, E.D.; Vidal, R.;“ Structured Low-Rank Matrix Factorization: Optimality, Algorithm, and Applications to Image Processing,” Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland USA, pp.4108-4117, 2014.

كشف و تجزئة الاجسام البارزة بالاعتماد على خوارزمية تباين المنطقة

ا. م. د. مثيل عمادالدين عبدالمنعم*

المستخلص: تعد عملية كشف الأجسام البارزة من المهام الرئيسية المهمة في مجال الرؤية الحاسوبية والتي تهدف الى كشف وجود الأجسام المهمة في الصورة الثابتة او في الصور المتسلسة (الفيديو) او ايجاد الهدف المقصود في المشهد. ان كشف الأجسام البارزة لها اهميه كبيرة في العديد من المجالات مثل تمييز الأجسام، انظمة المراقبة، توضيح واسترجاع الصور، وغيرها. في هذا العمل تم اعداد طريقة لكشف وتقطيع الاجسام البارزة من الصور والفديو في حالة كان الهدف متحرك او متوقف. بالمقارنة مع طرق كشف الاجسام البارزة التقليدية مثل تقنية طرح الخلفية فأن نموذجنا المقترح يتمكن من كشف الأجسام البارزة في حالة كان الجسم متحرك او ثابت ويتمكن من فصل الجسم البارز من الخلفية مع تفاصيل عالية بنسبة كشف تساوي 98.05.

الكلمات المفتاحية: تجزئة الصورة، الشيء الامامي، طرح الخلفية.

* الجامعة التكنولوجية، بغداد، العراق