Published in Conference Computer Analysis of Images and Patterns
(In: Lecture Notes in Computer Science) 2124, 630-637, 2001
which should be used for any reference to this work

Visual Attention Guided Seed Selection for Color Image Segmentation

Nabil Ouerhani[1], Neculai Archip[2], Heinz Hügli[1], and Pierre-Jean Erard[2]

[1] Institute of Microtechnology, University of Neuchâtel
Rue A.-L. Breguet 2, CH-2000 Neuchâtel, Switzerland
{Nabil.Ouerhani, Heinz.Hugli}@unine.ch
[2] Institute of Computer Sciences, University of Neuchâtel
Emile-Argand 11, CH-2000 Neuchâtel, Switzerland
{Neculai.Archip, Pierre-Jean.Erard}@unine.ch

Abstract. The "seeded region growing" (SRG) is a segmentation technique which performs an image segmentation with respect to a set of initial points, known as seeds. Given a set of seeds, SRG then grows the regions around each seed, based on the conventional region growing postulate of similarity of pixels within regions. The choice of the seeds is considered as one of the key steps on which the performance of the SRG technique depends. Thus, numerous knowledge-based and pure data-driven techniques have been already proposed to select these seeds. This paper studies the usefulness of visual attention in the seed selection process for color image segmentation. The considered purely data-driven visual attention model provides the required points of attention which are then used as seeds in a SRG segmentation algorithm using a color homogeneity criterion. A first part of this paper is devoted to the presentation of the multicue saliency-based visual attention model, which detects the most salient parts of a given scene. The possibility of using the so far detected regions as seeds to achieve the region growing task is then discussed. The last part is dedicated to experiments involving a variety of color images.

1 Introduction

Visual attention is the ability to rapidly detect interesting parts of a given scene. Using visual attention in a computer vision system permits a rapid selection of a subset of the available sensory information before further processing. The selected locations are supposed to represent the conspicuous parts of the scene. Higher level computer vision tasks can then focus on these locations.
Various computational models of visual attention have been presented in previous works [1–3]. These models are, in general, data-driven and based on the feature integration principle [4]. Known as saliency-based, the model presented in [1] considers a variety of scene features (intensity, orientation and color) to compute a set of conspicuity maps which are then combined into the final saliency map. The conspicuity operator is a kind of "contrast detector" which, applied on

a feature map, detects the regions of the scene containing relevant information. Visual attention processes have been used to speed up some tasks, for instance object recognition [5], landmarks detection for robot navigation [6] and 3D scene analysis [7]. The image segmentation task considered further should also benefit from visual attention.

The "seeded region growing" (SRG) presented in [8] is a segmentation technique which performs a segmentation of an image with respect to a set of points, known as seeds. SRG is based on the conventional region growing postulate of similarity of pixels within regions. Given a set of seeds, SRG then finds a tessellation of the image into homogeneous regions. Each of which is grown around one of the seeds.

It is obvious that the performance of SRG technique depends strongly on the choice of the seeds. Some previous works have dealt with the seed selection problem [9]. Knowledge-based as well as pure data-driven solutions have been proposed. The first class of methods is usually used in specific contexts where information about the regions of interest is available. Automatic knowledge based methods as well as pure interactive seed selection belong to this class. The data-driven methods are, however, more general and can be applied on scene images without any a priori knowledge. Consequently, a wider range of images can be processed using the latter class of techniques. Numerous data-driven seed selection methods are based on histogram analysis [10]. Using either original images or even gradient images, the technique aims to find peaks on the histogram. These peaks are supposed to constitute homogeneous regions on the image. A suitable thresholding permits the selection of these regions. The selected locations of the image are then used as seed regions to achieve the SRG task. This seed selection method is straightforward for gray level images and can be extended to deal with color images. A seed selection based on intensity and color needs, however, a mechanism which combines both features.

In this work we study the possibility to use visual attention as a method for seed selection from color images. This idea is motivated by the performance of the bottom-up saliency-based model of visual attention to detect interesting locations of an image, taking into account a variety of scene features. The automatically detected salient regions are used as seeds to apply the SRG technique on color images. A first part of this paper is devoted to the presentation of the visual attention model used in this work. The SRG algorithm is then described. The last part of the paper is devoted to experimental results carried out on various color images.

2 Visual attention model

2.1 Saliency-based model

According to a generally admitted model of visual perception [2], a visual attention task can be achieved in three main steps (see Fig. 1).

1) First, a number (n) of features are extracted from the scene by computing the so-called feature maps. Such a map represents the image of the scene, based on

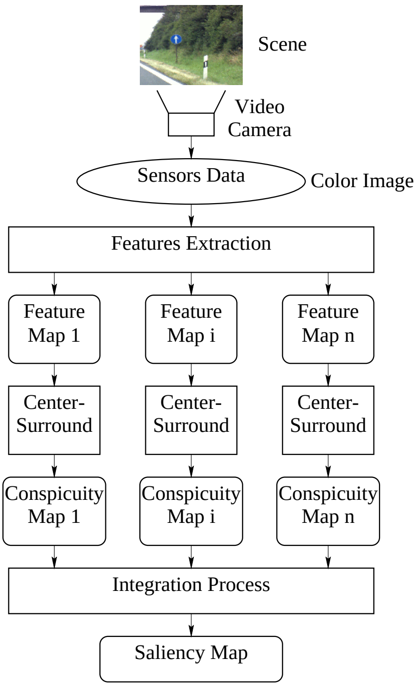

Fig. 1. Scheme of a computational model of attention.

a well-defined feature. This leads to a multi-feature representation of the scene. The features used in this work are intensity, color components, and gradient orientation.
2) In a second step, each feature map is transformed in its conspicuity map. Each conspicuity map highlights the parts of the scene that strongly differ, according to a specific feature, from its surrounding. In biologically plausible models, this is usually achieved by using a *center-surround*-mechanism. Practically, this mechanism can be implemented with a *difference-of-Gaussians*-filter, which can be applied on feature maps to extract local activities for each feature type.
3) In the last stage of the attention model, the n conspicuity maps are integrated together, in a competitive way, into a *saliency map* $\mathcal{S}$ in accordance with equation 1.

$$\mathcal{S} = \sum_{i=1}^{n} w_i \mathcal{C}_i \tag{1}$$

The competition between conspicuity maps is usually established by selecting weights w_i according to a weighting function w, like the one presented in [1]: $w = (M - \overline{m})^2$, where M is the maximum activity of the conspicuity map and $\overline{m}$ is the average of all its local maxima. w measures how the most active locations differ from the average. Thus, this weighting function promotes conspicuity maps in which a small number of strong peaks of activity is present. Maps that contain numerous comparable peak responses are demoted. It is obvious that this competitive mechanism is purely data-driven and does not require any a priori knowledge about the analyzed scene.

2.2 Selection of salient locations

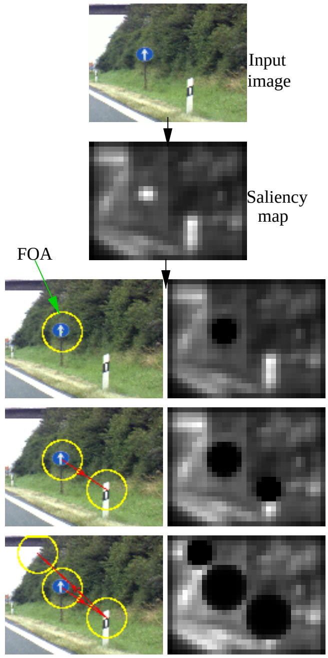

Fig. 2. Salient regions selection. Applying a WTA mechanism to a saliency map permits the selection of the most salient locations of the image.

At any given time, the maximum of the saliency map defines the most salient location, to which the focus of attention (FOA) should be directed. A "winner-take-all" (WTA) mechanism [1] is used to detect, successively, the significant regions. Given a saliency map computed by the saliency-based model of visual attention, the WTA mechanism starts with selecting the location with the maximum value of the map. This selected region is considered as the most salient part of the image (winner). The FOA is then shifted to this location. Local inhibition is activated in the saliency map, in an area around the actual FOA. This yields dynamical shifts of the FOA by allowing the next most salient location to subsequently become the winner. Besides, the inhibition mechanism prevents the FOA from returning to a previously attended locations. An example of salient regions selection based on the WTA mechanism is given in Figure 2.

3 Seeded region growing

Given a number of seeds, the SRG algorithm finds homogeneous regions around these points. Originally [8] this method is presented and applied to gray-scale images. We extend this algorithm for the color images. On the original image, we compute the FOA points (see Fig. 2) and we use them as seeds input for the region growing algorithm. Thus, for each spot we obtain a region. The algorithm we use to grow a region is:

Seeded algorithm

```
decompose original image on the R, G, B channels
create R(i), initial region from the seed point
add the neighborhoods  of R(i) in SSL
begin
repeat
  remove first point x from SSL
  if(x satisfy a membership criteria in R(i))
    add x to R(i)
  end if
  if(x was added in R(i))
    add into SSL the neighbors of x which are not in SSL
    update the mean of the region R(i)
  end if
until SSL is not empty
```

By SSL we denote a list with candidate points called sequentially sorted list. Initially it contains the neighborhoods of the seed pixel. To decide the criterion homogeneity of the regions, initially we compute the $Tolerance(I_R)$, $Tolerance(I_G)$, $Tolerance(I_B)$ the tolerance rates for the three channels, using the cluster decomposition of histograms. Let f be the function which denotes the image, and f_R, f_G, f_B the image functions for the three channels. We express:

$$\delta_R(x) = |f_R(x) - mean_{y \in R_i}[f_R(y)]|$$
$$\delta_G(x) = |f_G(x) - mean_{y \in R_i}[f_G(y)]|$$
$$\delta_B(x) = |f_B(x) - mean_{y \in R_i}[f_B(y)]|$$

Thus, the criteria that must be accomplished by the candidate point x is:

$\delta_R(x) < Tolerance(I_R)$ and
$\delta_G(x) < Tolerance(I_G)$ and
$\delta_B(x) < Tolerance(I_B)$

4 Experiments

Numerous experiments have been carried out in order to study the usefulness of visual attention for the seeded region growing algorithm. Four outdoor scenes have been considered in the experiments presented in Figure 3. Each scene is represented by its color image. A saliency map is computed for each color image, using the saliency-based model of visual attention presented in section 2.1. A winner-take-all (WTA) mechanism selects the eight most conspicuous parts of the scene from the computed saliency map (see section 2.2). The seeded region growing (SRG) presented in section 3 uses the selected locations as seed points. Eight regions are segmented, each of which is grown around one of the seeds.

The most salient point of the first scene is located on the traffic sign. This is due to color and intensity contrast in this part of the image. Starting the segmentation task at this location permits the segmentation of the arrow of the traffic sign. Due to intensity contrast two parts of the signpost are within the eight most salient locations. Through a targeted region growing around these two points, the main part of the signpost can be segmented. For the same reason, the fifth most salient location is situated on the road border line, which allows the segmentation of the whole road border. The part of the sky visible on the image (upper left corner) is segmented by applying the seeded region growing task around the third most salient location of the image. The largest segmented region represents the forest. It is grown around the fourth most salient point of the scene.

Eight salient locations are also selected from the saliency map computed from the color image of the second scene. Consequently, eight regions are segmented around these seeds. For instance, the arrows of the two blue traffic signs are segmented around the third and the sixth most salient positions of the image. The speed limitation sign contains the second most salient location. The segmentation around this point easily delimits the contour of the number '60'. Some parts of the car are also segmented.

In the third scene, two traffic signs are segmented, the first one indicating the directions and the second one containing speed limitations. Road borders as well as a part of the forest are also segmented. Three important traffic signs are segmented in the fourth scene. A part of the white car and a part of the discontinuous line separating two road lanes are also segmented.

It is important to notice that neither the visual attention model nor the segmentation algorithm are adapted to road scenes analysis. Nevertheless, important objects such as traffic signs or road border lines are often segmented. This kind of objects contain relevant information and often stand out from the rest of the scene, in order to be easily perceived by drivers. These characteristics are natural help to the artificial visual attention mechanism to detect these relevant scene elements.

The presented experiments clearly show the usefulness of a visual attention mechanism for color image segmentation by means of seeded region growing. The salient locations of the image are natural candidates of seeds since they are often surrounded by relevant information. An additional benefit of the combination of visual attention and the SRG algorithm is the speed up of the segmentation task.

5 Conclusion

This work studies the usefulness of visual attention in the seed selection process for color image segmentation. The considered purely data-driven visual attention process, built around the concepts of conspicuity and saliency maps, provides the required points of attention which are then used as seeds in a SRG segmentation algorithm using a color homogeneity criterion. The experiments presented in this paper clearly validate the idea of using salient locations as start points for the SRG algorithm. The experiments concern outdoor road traffic scenes. Despite the unavailability of any a priori knowledge about the analyzed scenes, the segmented objects include, in all cases, the most conspicuous road signs of the scene. The results speak for the good performance of the attention guided segmentation in similar scenes. Due to its bottom-up character, the reported segmentation method is expected to have similar performance in different scenes.

Acknowledgment

Part of this work was supported by the Swiss National Science Foundation under project number 54124.

References

1. L. Itti, Ch. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 20(11), pp. 1254-1259, 1998.
2. R. Milanese. Detecting salient regions in an image: from biological evidence to computer implementation. Ph.D. Thesis, Dept. of Computer Science, University of Geneva, Switzerland, Dec. 1993.
3. S.M. Culhane and J.K. Tsotsos. A prototype for data-driven visual attentional. ICPR92, Vol. 1, pp. 36-40, Hague, Netherland, 1992.
4. A.M. Treisman and G. Gelade. A feature-integration theory of attention. Cognitive Psychology, pp. 97-136, 12, 1980.
5. Aparna Lakshmi Ratan. The role of fixation and visual attention in object recognition. MIT AI-Technical Report 1529, 1995.
6. E. Todt and C. Torras. Detection of natural landmarks through multi-scale opponent features. ICPR 2000, Vol. 3, pp. 988-1001, Barcelona , September 2000.
7. N. Ouerhani and H. Hugli. Computing visual attention from scene depth. ICPR 2000, Vol. 1, pp. 375-378, Barcelona, Sep. 2000.
8. R. Adams and L. Bischof. Seeded region growing. IEEE Trans. on Pattern Analysis and Maschine Intelligence, vol 16, no 6, 1994.
9. U. Koethe. Primary image segmentation. 17. DAGM-Symposium, Springer, 1995.
10. L. O'Gorman and A.C. Sanderson. The converging squares algorithm: An efficient method for locating peaks in multidimensions. IEEE Trans. Pattern Analysis and Machine Intelligence, PAMI, Vol.6, pp. 280-288., 1984.

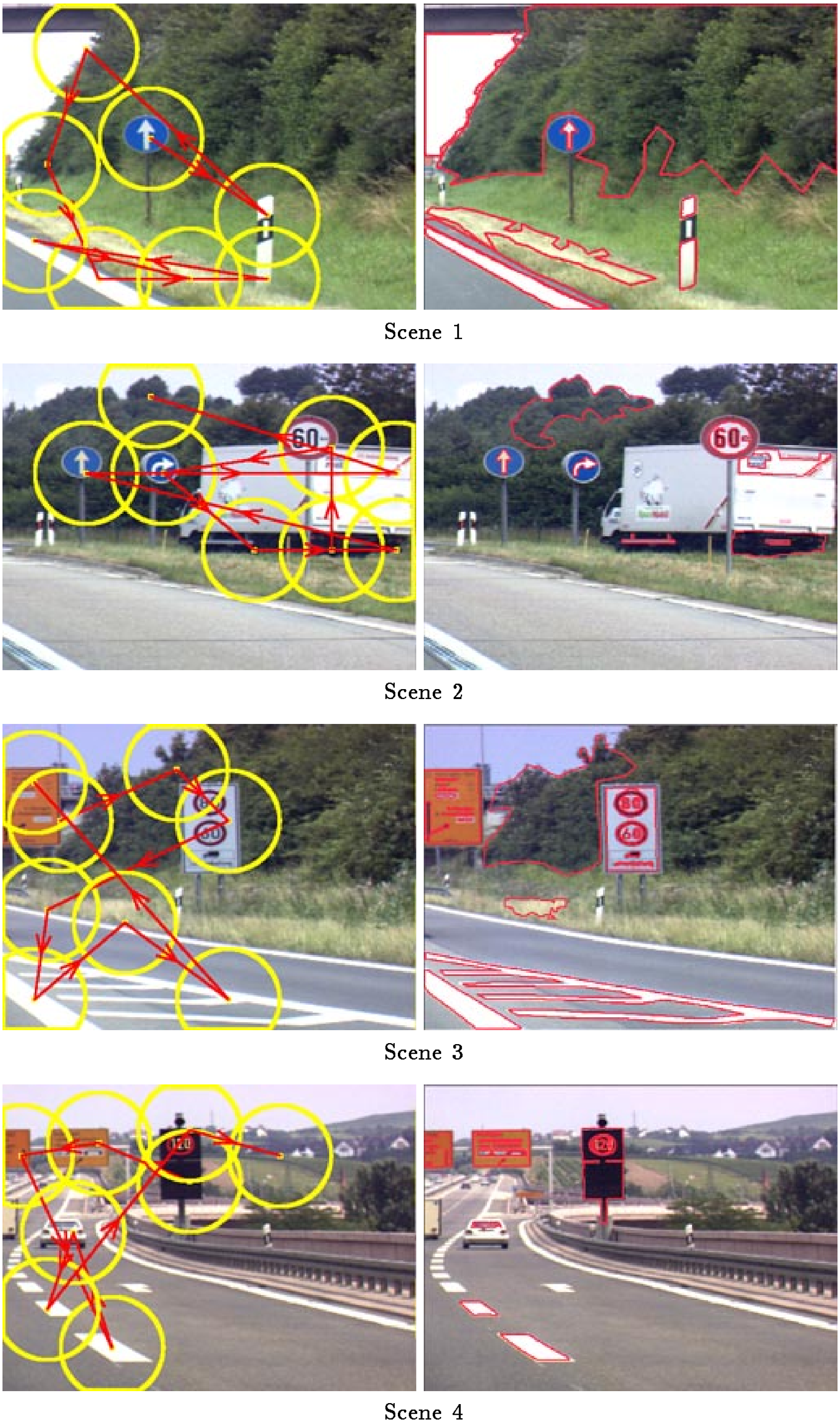

Fig. 3. Experimental results. Left: the color images with the eight most salient locations. Right: segmented salient regions.