A Color Image Segmentation Method Based on Seeded Region Growing and Visual Attention

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. This paper reports a color image segmentation method based on a seeded region growing technique (SRG) and guided by a saliency-based visual attention algorithm. Inspired by biological vision the purely data-driven model of visual attention is built around the feature, conspicuity and saliency maps. Using chromatic as well as unchromatic scene features, it, automatically, generates a set of regions-of-interest (ROIs), which represent the most visually-salient locations of the image. The automatically selected points are then used as seeds by the region growing algorithm to segment the conspicuous parts of the scene, using a color homogeneity criterion. A snakes-based technique is then used to improve the contours of the segmented regions. The results reported in this paper clearly show the effectiveness of the considered model of visual attention to detect the salient locations in color images. They also illustrate the usefulness of these automatically generated points to guide the seeded region growing algorithm.

1 Introduction

In biological vision, visual attention is usually used to refer to the ability of the visual system to rapidly detect "interesting" parts of a given scene. Inspired by biological vision, the principle of visual attention is used with a similar goal in artificial vision. Thus, using visual attention in a computer vision system permits a rapid selection of a subset of the available sensory information by generating a set of regions-of-interest (ROIs). The automatically selected locations are supposed to represent the most visually-salient parts of the scene on which higher level computer vision tasks can focus.
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] which is inspired by psychophysical studies on human visual attention. The buttom-up saliency-based model presented in [1] is built around the feature, conspicuity and saliency maps. Considering a variety of scene features (intensity, orientation and color), it computes a set of conspicuity maps, which are then combined into the final attention map, known as the

saliency map. The saliency map topographically encodes stimulus conspicuity, or saliency at every location of the scene.

The saliency-based model of visual attention has been successfully used to guide some computer vision-related tasks, such as object recognition [5], landmarks detection for robot navigation [6], 3D scene analysis [7] or even color image compression [8]. The image segmentation task considered further should also benefit from visual attention.

The "seeded region growing" (SRG) presented in [9] is a segmentation technique which performs a segmentation of an image with respect to a set of points, known as seeds. 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. Furthermore SRG is based on the conventional region growing postulate of similarity of pixels within regions.

It is obvious that the performance of SRG technique strongly depends on the choice of the seeds. Some automatic, histogram-based seed selection methods have been presented in previous works [10, 11]. These seed selection methods are 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 for the SRG technique. Furthermore the contours of the segmented regions are improved by means of a snakes-based technique.

This paper is organized as follows. Section 2 is devoted to the presentation of the visual attention model used in this work. Section 3 describes the seeded region growing algorithm. Section 4 reports the experimental results carried out on various color images in order to assess the effectiveness of visual attention in the segmentation process. Finally, the conclusions are stated in Section 5

2 Visual attention 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).

2.1 Feature maps

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 a well-defined feature. This leads to a multi-feature representation of the scene. This work considers the following features which are computed from an RGB color image.

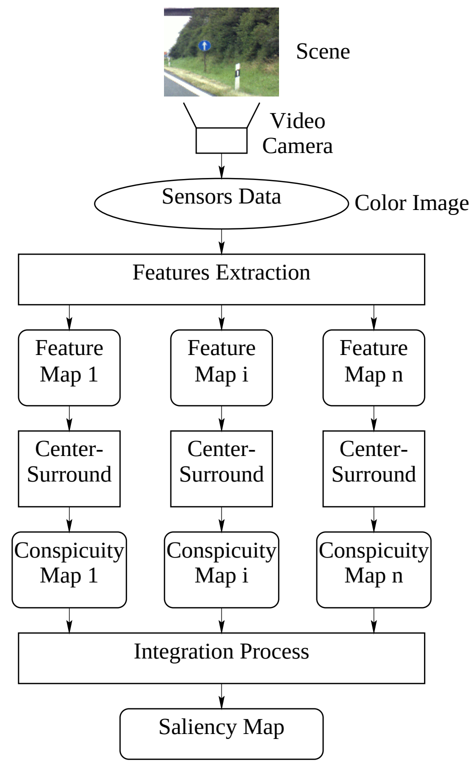

Fig. 1. Scheme of a computational model of attention.

- Intensity

$$I = (R + G + B)/3 \tag{1}$$

- Two chromatic features based on the two color opponency filters R^+G^- and B^+Y^- where the yellow signal is defined by $Y = \frac{R+G}{2}$. Such chromatic opponency exists in human visual cortex [12].

$$\begin{aligned} \mathcal{RG} &= R - G \\ \mathcal{BY} &= B - Y \end{aligned} \tag{2}$$

  Before computing these two features, the color components are first normalized by I in order to decouple hue from intensity.

2.2 Conspicuity maps

In a second step, each feature map is transformed in its conspicuity map which 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, $\mathcal{DoG}$ (See Eq.3), which can be applied on feature maps to extract local activities for each feature type.

$$\mathcal{DoG}(x,y) = \frac{1}{2\pi\sigma_{ex}^2} e^{-\frac{x^2+y^2}{2\sigma_{ex}^2}} - \frac{1}{2\pi\sigma_{inh}^2} e^{-\frac{x^2+y^2}{2\sigma_{inh}^2}} \tag{3}$$

A visual attention task has to detect interesting objects, regardless of their sizes. Thus, a multi-scale conspicuity operator is required. It has been shown in [2], that applying variable size center-surround filter on fixed size images, has a high computational cost. An interesting method to implement the *center-surround*-mechanism has been presented in [1]. This method is based on a multi-resolution representation of images. For each feature, nine spatial scales ([0..8]) are created using gaussian pyramids, which progressively lowpass filter and subsample the feature map. Center-Surround is then implemented as the difference between fine and coarse scales. The center is a pixel at scale $c \in \{2,3,4\}$ and the surround is the corresponding pixel at scale $s = c + \delta$ and $\delta \in \{3,4\}$. Consequently, six maps $\mathcal{F}(c,s)$ are computed for each pyramid $\mathcal{P}$ (**Eq. 4**).

$$\mathcal{F}(c,s) = |\mathcal{P}(c) - \mathcal{P}(s)| \tag{4}$$

The absolute value of the difference between the center and the surround allows the simultaneous computing of both sensitivities, dark center on bright surround and bright center on dark surround (red/green and green/red or blue/yellow and yellow/blue for color). A weighted sum of the six maps $\mathcal{F}(c,s)$ results into a unique conspicuity map for each pyramid and, consequently, for each feature. The maps are weighted by the same weighting function w described in Section 2.3.

2.3 Saliency map

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 5.

$$\mathcal{S} = \sum_{i=1}^{n} w_i \mathcal{C}_i \tag{5}$$

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.

2.4 Selection of salient locations

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 [13] 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 previously attended locations.

3 Image Segmentation

3.1 Seeded region growing

Define an image as a two-dimensional function $f(x, y)$ of the coordinates $x \in \{1, .., x_{max}\}$ and $y \in \{1, .., y_{max}\}$. Segmentation entails partitioning an image into regions that are coherent with respect to some criterion. The SRG algorithm finds homogeneous regions around a set of given points called seeds. Two problems are related with this method.
First one is the choice of the similarity criteria of pixels in regions. The second problem and more difficult is to select the initial seeds, which affect directly the quality of segmentation.
For the first problem, the similarity criteria, we choose to decompose the color image into the three channels R, G, B and to use the intensity on each of them as features which group together the pixels in regions. To decide about the membership of a point in a region, 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 three chromatic 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
$\delta_B(x) < Tolerance(I_B)$

For the second problem, the choice of the seed points, we use the spots obtained with the attention mechanism described previously, as seeds for SRG. Thus, for each spot we obtain a region. The algorithm used to perform the segmentation is the following:

Algorithm 1 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
while SSL is not empty do
  remove first point x from SSL
  if x satisfy a membership criteria in R_i then
    add x to R_i
  end if
  if x was added in R_i then
    add into SSL the neighbors of x which are not in SSL
    update the mean of the region R_i
  end if
end while
```

3.2 Snakes

Due to the presence of noise in the images, the quality of the contours is not optimal. To improve it, we are using a snakes-based algorithm. The snakes are introduced in the literature by Kass [14], and formulated as an energy-minimizing spline. Given the spline $v(s) = (x(s), y(s))$, he defined the energy function:

$$E_{total} = \int_0^1 E_{int}(v(s)) + E_{image}(v(s)) + E_{con}(v(s))\, ds \tag{6}$$

where E_{int} represents the internal energy of the spline, composed of a first-order term controlled by $\alpha(s)$, which makes the snake act like a membrane, and the second-order term controlled by $\beta(s)$, making the snake to act like a thin plate.

$$E_{int} = (\alpha(s)|v_s(s)|^2 + \beta(s)|v_{ss}(s)|^2)/2 \tag{7}$$

E_{image} is given by:

$$E_{image} = -|\nabla I(x, y)|^2 \tag{8}$$

so that the snake is attracted by the contours with large gradients.
Finally E_{con} gives rise to the external constraint forces given by the user.
The problems related with this algorithm are the initial position of the contours and the minimization procedure used, including the instability and a tendency for points to bunch up on a strong portion of an edge.

In our case, the first problem is solved. As initial position for the snakes, we are using the contours obtained with the SRG algorithm. The second problem is more complicated, and some authors have dealt with it [15], [16].

A Greedy algorithm is used, which allows a contour with controlled first and second order continuity to converge on an area of high image energy, in this case edges, with $O(nm)$ complexity, when n is the number of points in the contour and m is the size of the neighborhood in which a point can move during a single iteration.

Important is also to approximate as correct as possible the curvature. We are using the formulas:

$$\left|\frac{dv_i}{ds}\right|^2 \approx |v_i - v_{i-1}|^2 = (x_i - x_{i-1})^2 + (y_i - y_{i-1})^2$$

and

$$\left|\frac{d^2v_i}{ds^2}\right| \approx |v_{i-1} - 2v_i + v_{i+1}|^2 = (x_{i-1} - -2x_i + x_{i+1})^2 + (y_{i-1} - -2y_i + y_{i+1})^2$$

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 2. 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. A winner-take-all (WTA) mechanism selects the eight most conspicuous parts of the scene from the computed saliency map (see section 2.4). 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

A color image segmentation method based on seeded region growing and guided by visual attention has been reported in this paper. The saliency-based model of visual attention provides a limited set of image points, which represent the most visually-salient locations of the scene. These automatically generated points are used as seeds to guide the segmentation process, which allows the segmentation of the most conspicuous image regions. Visually irrelevant scene parts are not further processed by the segmentation task permitting the speed up of the process. The experiments reported in this work consider outdoor road traffic scenes. The results clearly validate the presented segmentation method. Despite the unavailability of any a priori knowledge about the analyzed scenes, the segmented image parts include, in all cases, the most visually-relevant objects.

Acknowledgment

This work was partially supported by the Swiss National Science Foundation under Grant FN-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 attention. 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. Proc. ICPR 2000, IEEE Computer Society Press, Vol. 1, pp. 375-378, Barcelona, Spain, Sep. 2000.
8. N. Ouerhani, J. Bracamonte, H. Hugli, M. Ansorge, and F. Pellandini. Adaptive color image compression based on visual attention. Proc. ICIAP2001, IEEE Computer Society Press, pp. 416-421, Palermo, Italy, Sep. 2001.
9. R. Adams and L. Bischof. Seeded region growing. IEEE Trans. on Pattern Analysis and Maschine Intelligence, vol 16, no 6, 1994.
10. U. Koethe. Primary image segmentation. 17. DAGM-Symposium, Springer, 1995.
11. 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.
12. S Engel, X. Zhang, and B. Wandell. Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature, Vol. 388, no. 6637, pp. 68-71*, Jul. 1997.
13. Ch. Koch and S. Ullman. Shifts in selective visual attention: Towards the underlying neural circuity. In L.M. Vaina (edt), Matters of Intelligence, pp. 115-141., 1987.
14. Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, pages 321–331, 1988.
15. A.A. Amini et al. Using dynamic programming for minimizing the energy of active contours in the presence of hard constraints. *Second International Conference on Computer Vision*, pages 95–99, 1988.
16. Donna J. Williams and Mubarak Shah. A fast algorithm for active contours and curvature estimation. *CVGIP: Image Understanding*, 55(1):14–26, 1992.

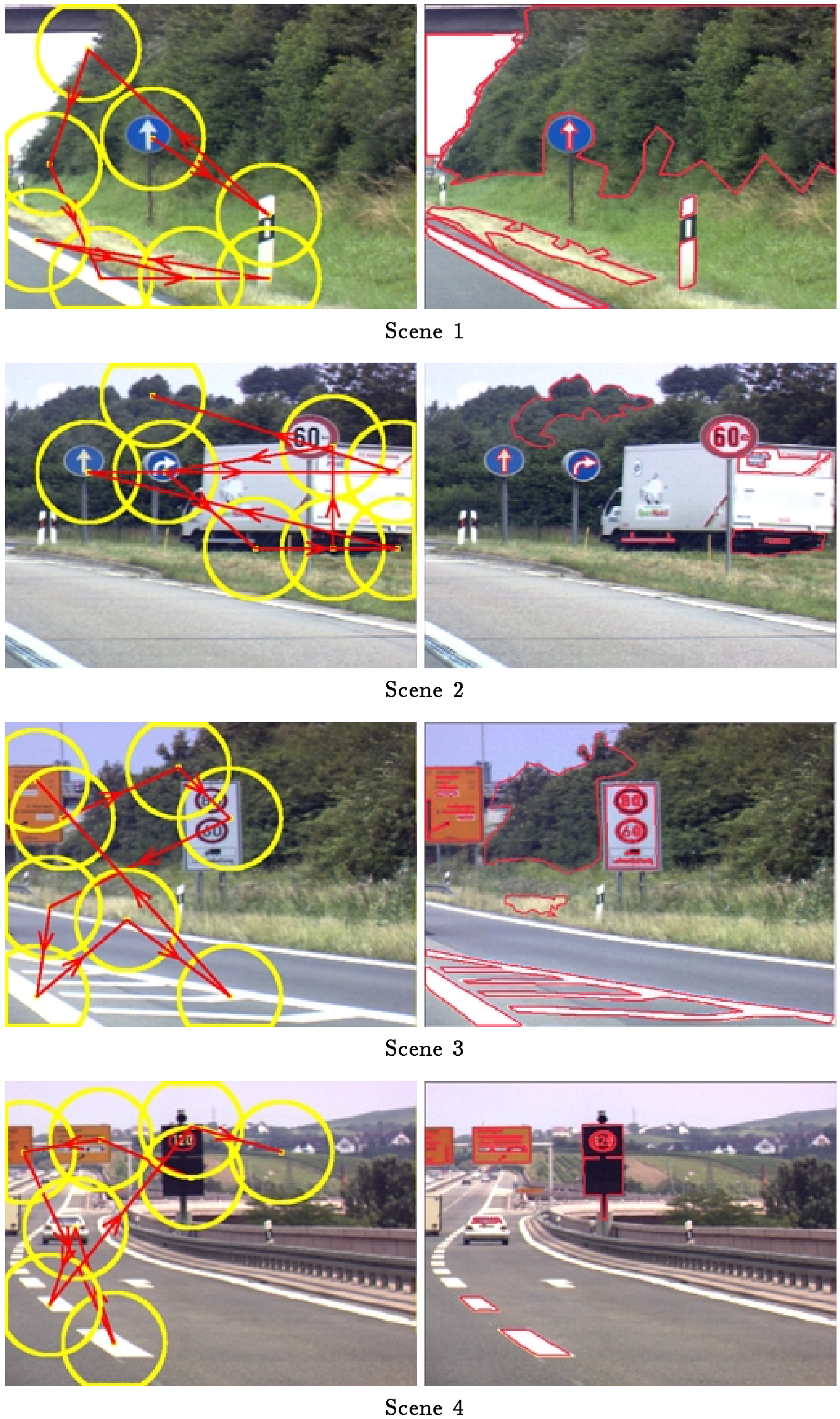

Fig. 2. Experimental results. Left: the color images with the eight most salient locations. Right: segmented salient regions.