A visual attention model for tracking regions based on color correlograms

LEANDRO LUÍS GALDINO, DÍBIO LEANDRO BORGES

Laboratório de Sistemas Inteligentes
Universidade Federal de Goiás, Escola de Engenharia Elétrica e de Computação
Praça Universitária sn, 74605-220, Goiânia, Go, Brasil
{leandro,dibio}@lis.eee.ufg.br

Abstract. The amount of information contained in a single color shot of a scene is extreme considering the variety of tasks that can be performed relying on visual data only. For the purpose of analyzing scenes dynamically, where objects come and go, one has to work under a structure of a visual attention model which priorizes what type of visual data the system has to follow. This paper presents a novel visual attention model for region tracking based on color correlograms. First, a reference frame is picked and it is segmented for the most significant color regions present in the shot. Color correlograms are then run on every frame after in order to provide spectral and spatial information, for the visual attention control, about the past and the new objects appearing on the scenes. The Visual Attention (VA) model then keeps track of old and new color regions appearing on the scene until a different frame is chosen. We have run experiments which show the performance of this proposed VA model.

1 Introduction

There are a variety of visual tasks that can not be performed considering the typical static model based paradigm known in Computer Vision. For the purpose of navigation and object identification through a sequence of frames for example, it becomes truly important to have an attention control module which helps to decide from frame to frame what and where to look next. Visual Attention (VA) models have appeared in the Computer Vision literature mainly after the coinning of terms such as Animate Vision [3], Active Vision [1], and Active Perception [2], where basically all of those terms refer to a change in paradigm where new models of analysis of scenes have to be designed considering Vision as being a dynamic process of information gathering and reasoning rather than an static picture by picture problem. The VA models commonly are proposed for filtering special geometric features such as corners and arcs from a scene and then deciding how those relate to the target situation.

One particular problem is to have a Visual Attention model segmenting and deciding upon regions to be tracked based on color information. Color is known to be a very important feature used for identifying objects in a scene, and specially under an Active Vision paradigm since the decision about what and where to look next for a visual task has to be made with an equilibrium of speed and precision, i.e. to make a correct choice under high uncertainty circunstances.

In this paper we propose a novel VA model which works segmenting a scene based on color information, and decides using color correlograms the most important regions to be tracked based on the initial reference frame. Two types of regions are tracked, first the most significant ones segmented from the reference frame, and then a novel region which is based upon what has not appeared in the first frame, i.e. what it is the most significant unknown region appearing in the sequence. The model shows promising results, and it could be used for a variety of tasks under the Active Vision paradigm. In the following section we present a review of the literature, give details of the VA model proposed by us, show some experiments, and draw conclusions about its performance and possible extensions.

2 Related Literature

Visual Attention models have been proposed in the literature for many purposes. They usually work integrating visual features in order to perform selection of candidate regions for posterior processing.

Itti, Koch, and Neibu [6] proposes an architecture where color, intensity and orientation features are extracted and a saliency map for each feature is built. For each map a center surround difference filter is applied in order to select a combination of winning features for pointing the next region of attention.

Gever and Smeulders [4] present a method for identifying objects inspired in the VA model by Treisman. This VA model has maps for selecting each visual feature, and a set of strategies for combining the maps. The purpose is not to select and track under novelty appearances.

Swain in [10] proposes a backprojection histogram technique for selecting color objects in a scene. The objects and

0-7695-0878-2/00 $10.00 © 2000 IEEE

colors are predefined in the RGB cube. The technique does not seem to scale up to work with palettes of approximate colors, and the results are not given for moving objects. Ratan in [8] presents results on a VA model for integrating some visual cues in a unique selection model. His main purpose was to use such a selection process for cutting an interpretation tree search for object recognition. The system was not applied for tracking and exploratory conditions.

3 A Visual Attention Model for Tracking Regions based on Color Correlograms

The VA model proposed in this paper, see Figure 6, gives an strategy for controlling visual attention using color information about the regions. Our model works first by choosing a reference frame from a video sequence, and on this picture a region growing segmentation algorithm is run for extracting the most significant color regions. The spectral information for these color regions are then normalized as the following.

Each channel is normalized with respect to the total signal from all three channels, effectively projecting the RGB color space onto the unit plane given by:

$$\hat{r} + \hat{g} + \hat{b} = 1 \tag{1}$$

Normalized RGB components are then computed as follows:

$$\hat{r} = \frac{r}{I} \tag{2}$$

$$\hat{g} = \frac{g}{I} \tag{3}$$

$$\hat{b} = \frac{b}{I} \tag{4}$$

$$I = \sqrt{\hat{r^2} + \hat{g^2} + \hat{b^2}} \tag{5}$$

In order to obtain the spectral information of the reference frame a color palette is organized in the following way, first the normalized [0,1] segment is divided into 16 equal bands, see Equation 1, and triples of the 3 channels we will have 4096 (16x16x16) buckets of the values in this palette. This particular number of bands was chosen for providing both color definition and speed processing for the task aimed. As it will be shown with the experiments results were successfull for frame rates above 15 per second. All values of the reference frame are indexed on this palette as seen in Equation 9 in order to find the spatial position of the pixel into the color palette. The values are referred as in Equations 6, 7, and 8, where *DIV* is a function for extracting the integer part of a number.

$$R = DIV(\hat{r} \times 15) \tag{6}$$

$$G = DIV(\hat{g} \times 15) \tag{7}$$

$$B = DIV(\hat{b} \times 15) \tag{8}$$

$$POS = \sum (R \times 16^2), (G \times 16^1), (B \times 16^0) \tag{9}$$

After having indexed all pixels in the color palette a search is run to verify positions where there were no pixels showing, and from this information we can verify what novelty values could be expected in the next pictures of the scene sequence. This allows a fast and efficient method for monitoring the appearance of new regions in the sequence, which is one of the aspects of the VA model for regions tracking.

Having the spectral information of the most significant regions, we will find them in the following frames by computing specific color correlograms of each of them as the main filters for signaling the VA model for tracking. A new region is expected frame by frame, and after it becomes of relevant size a color correlogram for it is also computed for the VA model to track.

In order to compute the color correlogram we will define a set of 16 distances, or 16x16x16 cubes in the RGB space, with the center at R=0, G=0, B=0. This way the first position, i.e. 1, in the color correlogram is the one which refers to the smallest distances between the colors, being the one which gives the exact presence of the region for tracking. The actual values of the color correlogram are computed using an euclidean distance of the RGB components of the pixels from each frame being processed.

Euclidean distance is a commonly used metric for comparing feature vectors in RGB space. A color correlogram express how the spatial correlation of color changes with the distance. Other methods such as color histogram for example captures only the color distribution in an image, whereas it does not give any spatial correlation information. The extended color correlogram method proposed here can be seen as one spatial extension of color histograms.

In the following equation I is defined as a frame in the sequence, its RGB components are I_r, I_g, and I_b, respectively. A pixel belongs to the first position of the color correlogram if, and only if,

$$H_1 = \sqrt{(I_r - R)^2 + (I_g - G)^2 + (I_b - B)^2} \quad (10)$$

$$F_1 = if H_1 < 16 \times \sqrt{3} \quad (11)$$

where R,G, and B are the pixel values of the color region in the reference frame.

4 Experiments

For the first experiments we have used a sequence containing 32 pictures, or 32 frames, with RGB of 24 bits, 60x80 pixels. Results given here show a subset of 6 of those frames. The images chosen were taken in the Laboratory for Intelligent Systems, using a digital camera under normal conditions of lighting. The images have stable conditions for the experiments, and an object being moved, which appeared after the reference frame was segmented.

Figure 6 shows a functional diagram of our approach. It starts by capturing the reference frame, with a specified number of regions initially set for segmentation. The spectral information of these initial regions are computed and a color correlogram is set for each region in order to find its correlate region in the next frames. The output of these feeds the VA model for tracking the regions, and a novelty region is expected after the reference frame. When this novelty region is detected, a color correlogram is set for it and it begins to be tracked from there on.

The reference frame used in these experiments is shown in Figure 5, R_1,R_2, and R_3 were specified as the regions to be tracked by the VA model. Any other regions could be specified, this particular set identify significant regions from this sequence. In Figure 1 it is shown the result of the regions segmentation by the color correlograms. The R_1 region, basically the background was well segmented, the same can be said for the R_3 region, for the region R_2 it can be noticed spots inside the region not segmented where this has occurred because of strong lighting differences on the face of the person. These spots though did not interfere in our results, since the model tracks more than 95% of the proposed regions. In Figure 2 the next frame in the sequence is shown, where regions R_1, R_2, and R_3 are well segmented, and it can be noticed a new region present in the scene. After a color correlogram is computed for the regions it still shows a white part in Figure 3 because the new region is not yet significant for the color correlogram. Figures 4, 7, 8, 9, 10, and 11 show the sequence of the four regions be tracked, the R_1, R_2, and R_3, and the new region shown in a separate window in black.

For the second set of experiments we have used a sequence containing 115 frames, with RGB of 24 bits, 96x72 pixels each frame. Results given here show a subset of 10 of those frames split into two sets. The first set contains 5 frames and the second also 5. This sequence has two new objects moving into and out of the scene after the reference frame was segmented.

Figures 17, 18, 19, 20, and 21 show results for the second set of experiments. The illuminance conditions interfere on the color of the objects, however in our experiments specially because of the robustness of the color correlogram approach the results were not altered. This has to happen in a practical system since for a exploratory navigation the environment conditions are unknown a priori. The combination of the VA model with the color correlogram computation for the regions shows to be precise and robust for the purpose of tracking regions based on color, since movement and change of ilummination do not foul the approach.

5 Conclusions

Results shown in this paper demonstrate that tracking objects, or regions can be done efficiently, even if color is the only visual characteristic used. The VA model presented here is an important tool for visual systems of exploratory navigation, where what is known and what is unknown are important elements to be analyzed and possibly tracked. We tested the approach under normal circunstances, and the results show precision and robustness. The VA model could be easily changed for triggering and tracking under different conditions such as by picking different reference frames along the sequence, and allowing for an increasing number of regions or objects appearing in the scene. So the model could be configured for many different situations, under the same approach.

Under the paradigm of Active Vision VA models are important elements for performing a great variety of tasks. The VA model presented here is novel and shows a promising line of research to be pursued for building up Active Vision systems for exploratory navigation, and robotic handling of objects by visual data input only. Color correlograms are robust solutions for finding uncertain information with good performance, i.e. speed and accuracy. The sequence shown here is segmented and tracked in real time running in a personal computer of 300 MHz clock.

As future work we plan to build up in the VA model by adding other visual features, such as texture and possibly geometric ones for performing 3D recognition.

Acknowledgements

We thank all the people from the Laboratory for Intelligent Systems, School of Electrical and Computer Engineering at Federal University of Goiás for the friendly environment, and especially Vivian B. Lasca for the frame sequence appearing in this paper.

References

[1] J. Aloimonos, I. Weiss, and A. Bandopadhay, Active Vision, *I.J. Computer Vision*, vol.1, pp.333–356, 1988.

[2] R. Bajcsy, M. Campos. Active and Exploratory Perception, *CVGIP: Image Understanding*, vol.56, n.1, pp.31–40, 1992.

[3] D. H. Ballard. Animate Vision, *Artificial Intelligence Journal*, vol.48, pp.57–86, 1991.

[4] T. Gevers and A. W. M. Smeulders. Image Indexing using Composite Color and Shape Invariant Features, *Proc. of the International Conference on Computer Vision*, Bombay, India, 1998.

[5] J. Huang, S. R. Kumar, M. Mitra, W. Zhu, and R. Zabih. Spatial Color Indexing and Applications, *I.J. Computer Vision*, vol. 35, n.3, pp.245–268, 1999.

[6] L. Itti, C. Koch and E. Niebur. A Model of Saliency-based Visual Attention for Rapid Scene Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* ,vol.20, n.11, pp.1254-1259, 1998.

[7] L. Itti and C. Koch. Learning to detect salient objects in natural scenes using visual attention, *Proceedings of DARPA 98*, USA, pp. 1201–1206, 1998.

[8] A. L. Ratan. *The Role of Fixation and Visual Attention in Object Recognition*, M.Sc. Thesis, Department of Electrical Engineering and Computer Science, MIT, USA, July 1995.

[9] K. Sung. *A Vector signal Processing Approach to Color*, M.Sc. Thesis, Department of Eletrical Engineering and Computer Science, MIT, USA, January 1992.

[10] M. J. Swain. *Color Indexing*, Ph.D. Thesis, Department of Computer Science, University of Rochester, USA, November 1990.

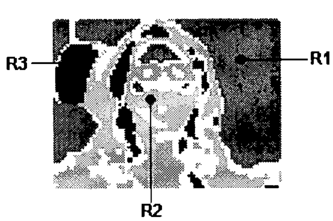

Figure 1: The result of region segmentation before running the color correlogram.

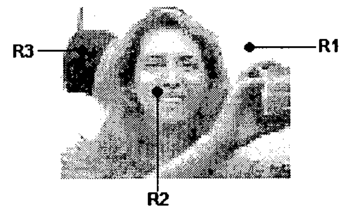

Figure 2: The regions defined in the first frame found in the second frame

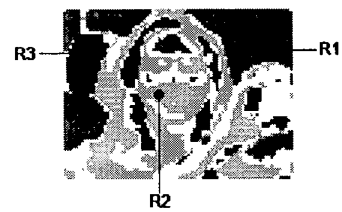

Figure 3: Segmented second frame

Figure 4: Detected new region

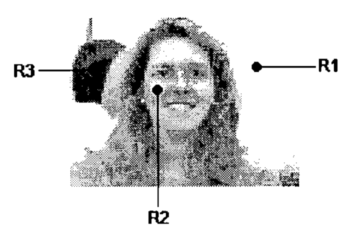

Figure 5: Original image, here we defined the most important regions for the color correlograms, which are passed for tracking in the next frame.

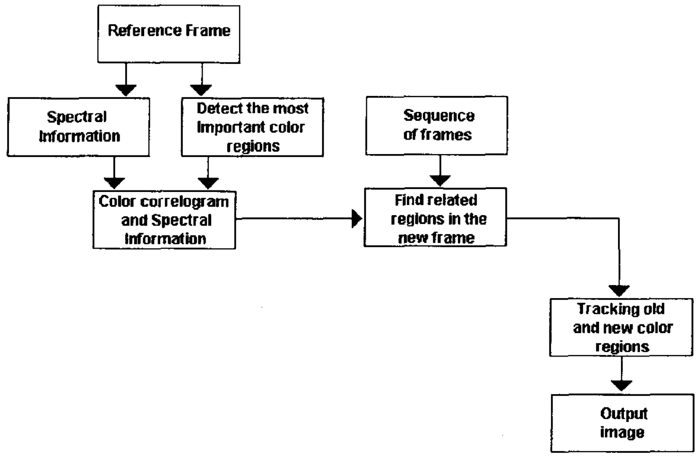

Figure 6: Visual Attention Model Diagram

Figure 7: Second image presented to the VA model, 3 regions tracked, and detected new region

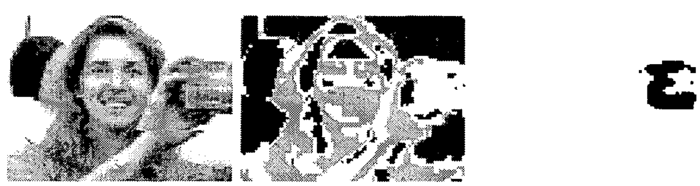

Figure 8: Third image presented to the VA model, 3 regions tracked, and detected new region

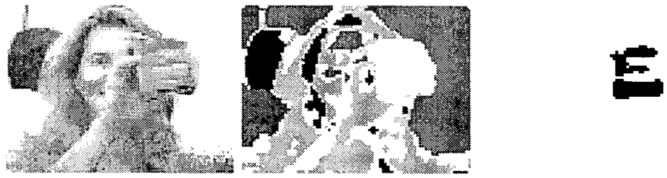

Figure 9: Fourth image presented to the VA model, 3 regions tracked, and detected new region

Figure 10: Fifth image presented to the VA model, 3 regions tracked, and detected new region

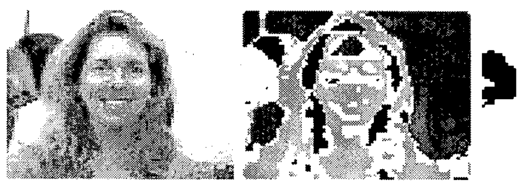

Figure 11: Sixth image presented to the VA model, 3 regions tracked, and detected new region

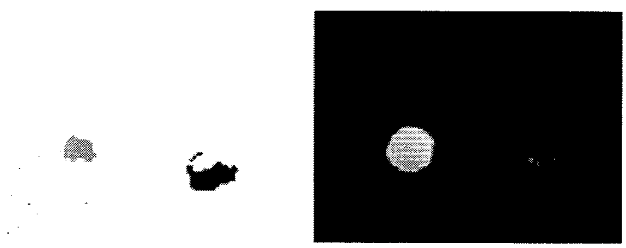

Figure 12: First frame presented to the VA model, the objects detected and tracked.

Figure 13: Second frame presented to the VA model, 2 regions detected and tracked, and the new one also detected and tracked.

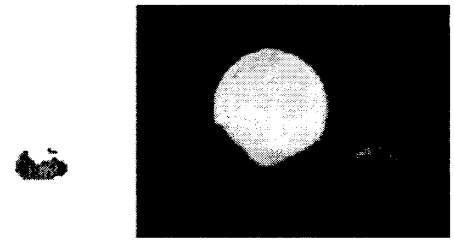

Figure 14: Third frame, 2 regions detected and tracked.

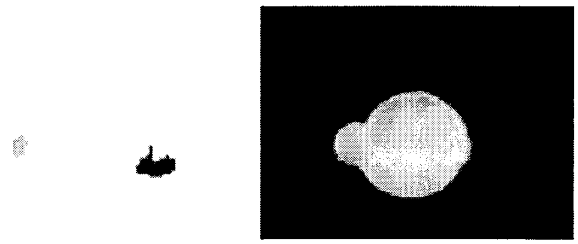

Figure 15: Fourth frame, 3 regions detected and tracked.

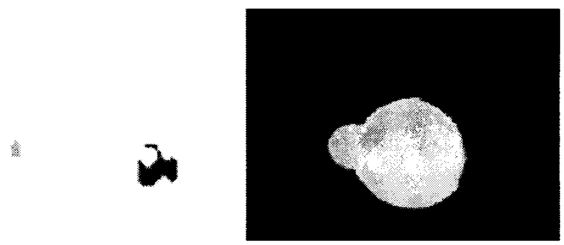

Figure 16: Fifth frame, 3 regions detected and tracked.

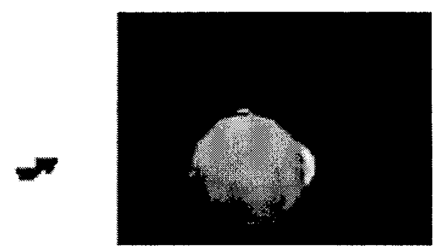

Figure 17: First frame of the second set of frames, 2 regions detected and tracked.

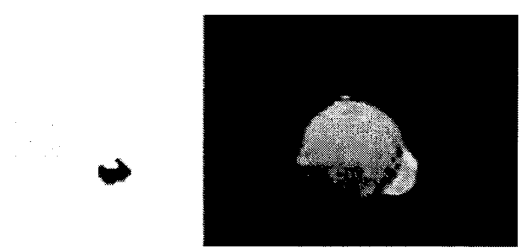

Figure 18: Second frame, 2 regions detected and tracked.

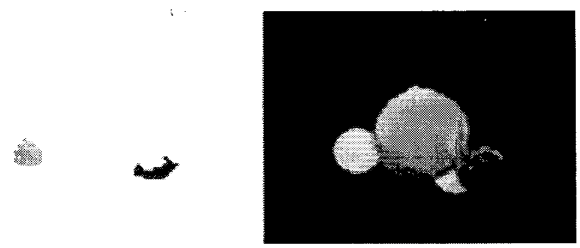

Figure 19: Third frame, 3 regions detected and tracked.

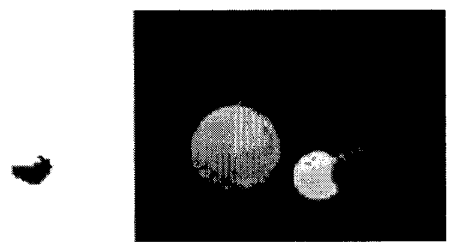

Figure 20: Fourth frame presented to VA model, 2 regions detected and tracked.

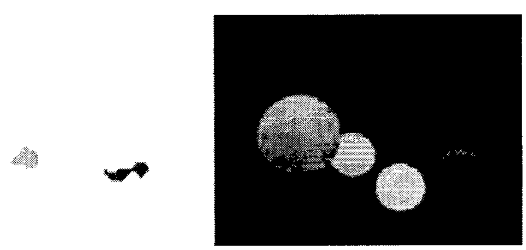

Figure 21: The last of the second set of frames, 3 regions detected and tracked.