GAZE: An Attentional Mechanism for the Extraction of Prominent Facial Features

R. Herpers, H. Kattner and G. Sommer*

GSF - Medis Institut
Institut für medizinische Informatik und Systemforschung
Ingolstädter Landstr. 1 D-85764 Oberschleißheim
Email: herpers@gsf.de

Abstract

In this paper a technical realization of an attentional mechanism extracting salient regions in face images is presented. Using the knowledge of the eye movement strategies of the human visual system the attentional mechanism localizes prominent regions in natural images. Elementary visual features are derived from high-resolution digitized images and represented in a so called topographical attention map. A top-down algorithm analyzes this map implemented as a multi-scale representation by searching the most salient features in each represented scale. The scale information is propagated from coarse scales to the next higher resolutions while the extent of the fixated region is adapted. After the extraction the selected and now locally limited regions are analyzed and interpreted. The proposed attentional mechanism is part of an image processing system identifying faces with dysmorphic signs. To improve the classification of a local analysis module, it is intended to examine these characteristic areas with specialized algorithms. This knowledge based approach shows that an artificial attention control system is capable of localizing the prominent facial areas (eyes, mouth, nose, etc.) in face images only based on elementary image features. The proposed algorithm can also be used to detect other attentive regions in any images based on appropriate features.

1. Introduction

Image processing algorithms whose realization do not include attentional influences have been shown to be computationally intractable [1]. The visual search task in the bottom-up case is NP-Complete in the size of the image, while the task-directed case has linear time complexity in the number of items in the display. These complexity considerations suggest that attentional mechanisms may be required to perform successfully an image analysis problem in real time.

Inspired by the eye movement strategies of the human visual system [2] the attentional mechanism selects prominent areas of the image to analyze only the information essential to a current task [3]. Irrelevant image regions are ignored and do not need to be processed further. Thus, image processing with attention control simplifies computation and reduces the amount of processing [4].

*Present address: Institut für Informatik, Lehrstuhl Kognitive Systeme, Universität Kiel, Preußerstr. 1-9, D-24105 Kiel

At the early stage of preattentive processing elementary features like motion, color, orientation, edge information and others are derived in the human visual system [5]. These features establish a preattentive representation, the so called "primal sketch" [6]. The control of the visual attention is based on this representation, which localizes spatial regions of interest in the order of their importance.

This paper presents a technical realization of an attentional mechanism simulating these physiological strategies. The approach is based on investigations of J. Tsotsos, who first developed "a prototype for data-driven visual attention" [7, 8]. Certain aspects of this model are not addressed in the presented paper, such as the inhibition of not relevant regions or the neural like modelling. Instead, emphasis is placed on a fast data-driven realization of an attentional mechanism localizing the prominent regions in face images. Our attentional mechanism will be used as an essential component in the building of a complete image processing system identifying faces with dysmorphic abnormalities [9].

The proposed model consists of a topographical attention map coded in a multiscale representation and an algorithm propagating the selected regions from coarse scales to the next higher resolution. For the results presented in this paper, two of the elementary visual features, edge and brightness information, are derived to guide the selective attention. The potential of the attention scheme is proposed by using facial 512x512 digitized gray-level images of normal and dysmorphic children.

2. Methods

The proposed attentional mechanism consists of a set of hierarchical computations. Task-oriented image information is derived and coded in a topological attention map comparable to preattentive visual stimuli. This map is transformed to a multi-scale representation and in the following analyzed to detect attentive image regions. A localized area is adapted to the extent of the underlying visual object and is extracted. After a local analysis it is suspended from the following processing of the attentional mechanism and the next salient region can be computed.

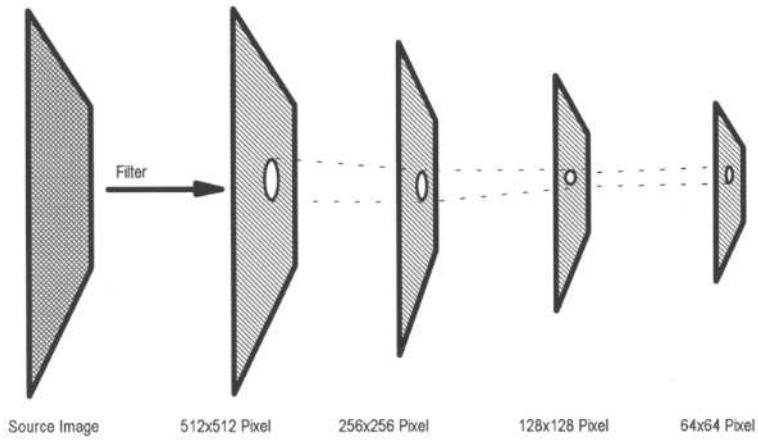

Fig. 1. Model of the multiscale representation. First, the original gray level image, the topological attention map in the highest resolution and 3 levels of an octave representation

2.1 Multi-scale representation

At the highest resolution, 512 x 512 pixel, the gray level image is transformed to a task-oriented image representation. The fast data-driven preattentive representation is given by the edge information provided by a 5 x 5 sobel filter, or alternatively by the brithtness information. The topological attention map is initialized with task-oriented weights which emphasize global image areas. For example, the image areas in the center are more important than the corners. The extracted image information is represented in the topological attention map by multiplication with a constant initialization. Multiresolution techniques are used to present the attention map in different scales [10]. We use a 5 x 5 gaussian convolution kernel to reduce the representation size (Fig. 1). The selection of an optimal scale representation is task-oriented and is one issue of this study. In a first approach tests were carried out using 4 or 5 octave pyramid levels and later 4 or less non octave levels of fixed size, because they enhance support of the detection of salient facial regions.

2.2 Adaptation of the extent

At the lowest resolution level of the attention map the actual, most attentive region is selected by a region based maximum search algorithm. The localized region is propagated through the entire hierarchy of the multi-resolution pyramid. The local extent of the underlying visual stimulus is adapted in a way that attentive visual objects will be detected completely. In each representation the region-growing is controlled by an algorithm which is determined by (1):

$$f(A) = 1 + \frac{A-1}{a\ (A-1) + b}; \qquad a, b > 0; \qquad (1)$$

where A is the area size and a, b are parameters determining the asymptotic properties of the region-growing algorithm [7,8].

In this study horizontal or vertical rectangles are used for matching attentive image regions (Fig. 2). During the adaptation process the edges of the calculated rectangle are expanded or reduced at each side of the rectangle.

2.3 Suspension of analyzed regions

At the highest resolution the detected regions are extracted and analyzed with specialized local algorithms [11]. After the analysis and local interpretation the focused regions are suspended from the following processing steps of the attentional mechanism. The intensity of the visual stimulus coded in the attention map is reduced to zero. We have examined two different procedures:

1) Constant reduction of the intensities in the whole detected area.
2) Gaussian like degradation beginning at the center of the extracted rectangle.

Schedular (2) produces a smooth alteration at the borders of the selected attentive region while the first one produces sharp changes. Very small attentive regions will suspend only corresponding small areas using the first proceeding. The second one

can be tuned to suspend larger regions. We also investigated a temporal delay which gradually reduces the suspension. The temporal degradation enables a repeated detection of formerly suspended regions. Applying this procedure a dynamical artificial eye movement record is created during a continued processing of attentive image regions (Fig. 3).

3. Results and discussion

We have shown that an artificial attention mechanism based on elementary features is able to detect salient regions in face images. In this paper, only brightness and edge information is used as input to the presented mechanism. These two basic features are sufficient for the detection of prominent facial regions, although the use of other features is possible.

The selected regions (Fig. 2) as well as the computed scan path (Fig. 3), which resembles the eye movement strategies observed in the human visual system, show an appropriate accordance with the salient image areas.

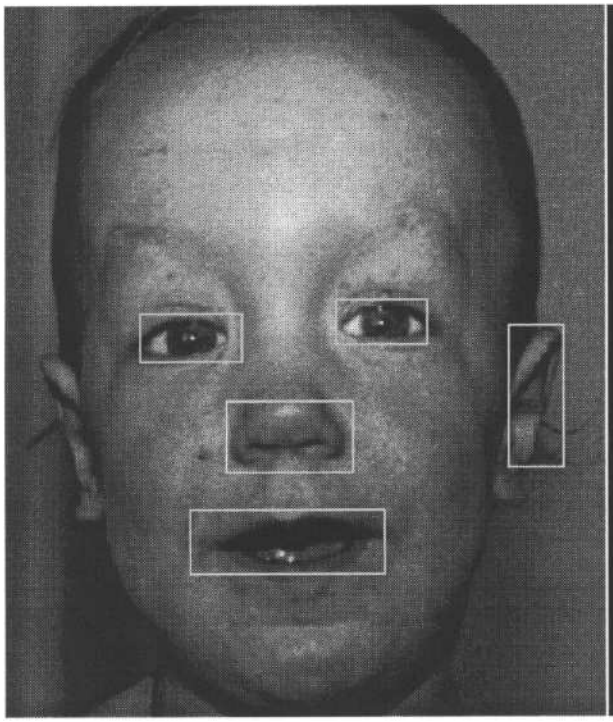

Fig. 2. Detection of characteristic facial regions

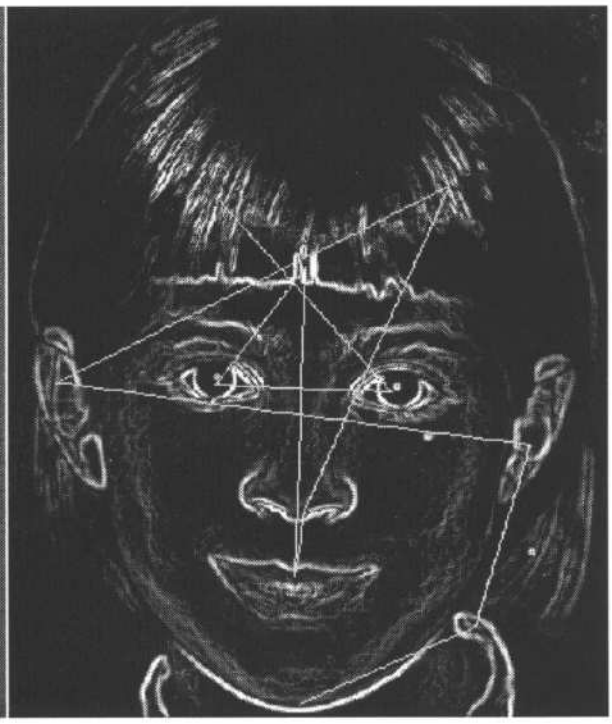

Fig. 3. Computer based attention record

The shown model is implemented in software on a SGI Indigo R4000 workstation and the measured performance in computing an attentional path with 20 steps lies between 7 and 10 seconds using high resolution, 8-bit, digitized, gray level images. The proposed algorithm is also tested on other real world images. The computed attentional mechanism localizes the salient areas in the natural scenes independent of the structure of the image data and the underlying task. There are no prerequisites to be taken like normalization or segmentation of the image data before the start of

computation. The presented model can handle all sizes and arrangements of image objects within the processed data. The use of the model of processing hierarchies combined with a maximum search as against a neural based approach with inhibitory qualities enhances the performance of a sequential processing of the attentional mechanism.

Acknowledgements

We appreciate the support of Prof. Dr. S. Stengel-Rutkowski and Dr. C. Apacik, Institut für Soziale Pädiatrie und Jugendmedizin, Abteilung Genetik der Ludwig-Maximilian-Universität München, providing the images used in this study. We also thank Dr. W. Eckstein, Institut für Informatik IX der Technischen Universität München for helpful suggestions and stimulating discussions.

References

[1] J. K. Tsotsos, *The complexity of perceptual search task,* In: Proc. of the IJCAI, Detroit, pp. 1571-1577 (1989)

[2] A. L. Yarbus, *Eye Movements and Vision,* New York: Plenum Press (1967)

[3] C. Koch and S. Ullman, *Shifts in selective visual attention: Towards the underlying neural circuity,* Human Neurobiol. Vol. 4, pp. 279-283 (1985)

[4] B. Olshausen, C. Anderson and D. v. Essen, *A neural model of visual attention and invariant pattern recognition,* Tec. Rep CNS Memo 18, Caltech Pasadena (1992)

[5] A. Treisman, *Preattentive processing in vision,* CVGIP Vol. 31, pp. 156-177 (1985)

[6] D. Marr, *Vision,* W. H. Freeman, San Francisco (1982)

[7] S. M. Culhane and J. K. Tsotsos, *A prototype for data-driven visual attention,* In G. Sandini (ed.) Computer Vision ECCV'92, Springer-Verlag, pp. 551-560 (1992)

[8] J. K. Tsotsos, *Localizing stimuli in a sensory field using an inhibitory attentional beam* Tec.-Rep. M5s 1A4, University of Toronto, Dep. of Comp. Sc., Canada (1991)

[9] R. Herpers, H. Rodax and G. Sommer, *A neural network identifies faces with morphological syndromes,* In this proceedings (1993)

[10] P. J. Burt, *Multiresolution techniques for image representation, analysis and smart transmission,* SPIE Conf. 1199, pp. 1-14 (1989)

[11] M. Michaelis and G. Sommer, *Characterization of key-points in images,* accepted by Int. Symp. on Subst. Ident. Techn. '93, Innsbruck (1993)