A Fuzzy Logic Model of Visual Importance for Efficient Image Synthesis

Ross Brown[a],
[a] Smart Devices Laboratory
Queensland University of Technology
2 George Street, Brisbane, Queensland, 4001, Australia.

Binh Pham[a], Anthony Maeder[b],
[b] School of Electrical and Electronic Systems Engineering, Queensland University of Technology, Australia.

Abstract

Experiments have shown that viewers only regard certain informative regions within a presented image. Furthermore, it has been shown that these regions of interest contain low level feature differences that influence the fixation time of the viewer. In this paper we present a novel fuzzy logic model of visual attention which seeks to compute the relative visual importance of regions in an image based upon these spatial feature differences. We also demonstrate some of the possible savings to be had in applying the visual importance model to the modulation of super-sampling in a ray-traced image. We expect this approach to have applications in the entertainment industry where image fidelity may be sacrificed for efficiency purposes.

I. Introduction

Spatial feature detectors in the *human visual system* are believed to interact in an early *preattentive* stage, causing regions of the viewing field to attract the attention of a viewer [1].

In a previous paper [2] we described a conceptual framework for the development of a novel feature-based scene decomposition system. This was based on the observation that there is a lack of a complete feature hierarchy for preattentive visual features [1]. Furthermore, we described preliminary psychophysical experiments performed during the design of the system [2]. We now extend this work into a more complete model of visual attention.

Our main goal for this model is its use in efficient adaptive rendering techniques in 3D computer graphics, in particular, the rendering technique known as *ray-tracing*. We believe that the application of a region-based visual attention model will allow rendering systems to judiciously apply resources to the regions being regarded by the viewer.

Most present adaptive image synthesis systems use image-space attributes to adaptively change the number of rays cast per pixel, mostly using statistical schemes [3]. One author has applied a multi-resolution visual attention model to the task of controlling caching and ray-tracing parameters in radiosity [4]. However, no region-based approaches have been used to determine the visual importance of image-space regions for image synthesis.

Preattentive feature-based models have been applied successfully to image processing, in particular, to the areas of active vision systems [5]. Other systems exist in the area of image compression that apply a visual importance model to a region segmentation using either a crisp computational model [6] or a fuzzy system [7]. These systems then heavily compress the less visually important regions of an image.

In our system, we seek to fuzzify difference thresholds, which are the basis for psychophysical models of visual attention [1]. Secondly, we seek to have a more complete model of feature integration based upon global image effects [8], including using the novel idea of contour densities to quantify the effect of high frequency information [9]. Lastly, we apply this region-based visual attention model in a novel manner to efficient image synthesis.

An adaptive image synthesis approach, incorporating a fuzzy visual attention model, facilitates image synthesis efficiency by applying the highest level of detail to the most perceptually salient regions of the scene. This visual feature model should have applications in other areas needing a region-based visual attention model.

Extending and applying previous work into new areas, we present in this paper the design, implementation and evaluation of a system that controls the sampling rate of a progressive ray-tracing system. The system allocates relative visual importance values to the regions using a fuzzy logic visual feature model. We describe the conceptual framework for the fuzzy system, and the design of the parameters used in the system. The membership functions and implication scheme for the importance model are detailed, followed by the integration of the fuzzy importance module into an adaptive rendering system. The paper then concludes with a description of future work to further extend the model and its applications.

II. System Conceptual Framework

The decomposition system is constructed from three major modules: a *sample generator*, a *quadtree image subdivision* and a *fuzzy logic visual importance module*, see Fig. 1.

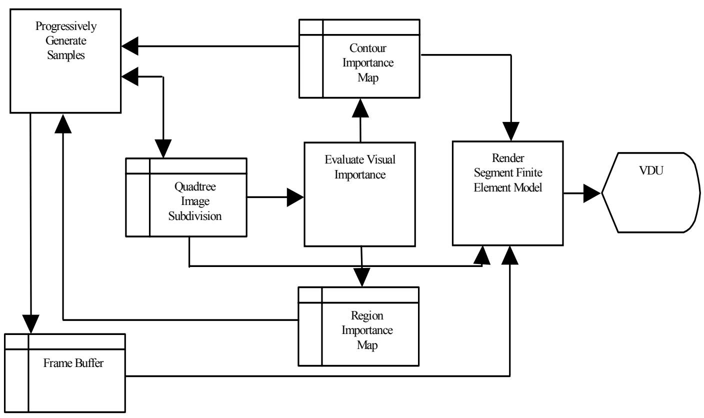

Fig. 1. Overview flow diagram of the attention-based ray-tracing system.

The sample generator controls the number of samples taken at any pixel location in the rendering of a scene. In order to overcome aliasing effects, a ray-tracing image synthesis system will sample a given area at a rate high enough to prevent jagged edges appearing in the image. We believe that this sampling rate can be reduced in areas that are not noticed by a viewer.

The sample generator recursively segments an image into quadrants based upon the visual importance of a region gained from a fuzzy logic contour importance model, storing this information in a quadtree subdivision. The quadtree representation is adaptive in nature, with the segmentation reflecting the objects contained within the scene.

Theoretically, the basis for this work is drawn from the principle that the visual saliency of a region is modelled by the mean feature value difference between a segmented region and its local neighborhood. In addition, it has been shown that global scene feature variability enhances or suppresses the previously mentioned saliency [8]. That is, the visual salience of the target is influenced by the surrounding noise. These global effects are modelled using the mean value of the feature differences across the entire scene.

Once the scene has been refined to an acceptable level, the image needs to be super sampled in order to provide the final high quality image. At this stage, a region-based visual attention model is applied to guide the super-sampling of each pixel. The fuzzy logic model assesses the visual importance of the image regions, integrating both local and global scene statistics. The model calculates the importance of these visual regions, which then biases the error tolerance parameter for any stop condition on the image refinement.

III. Fuzzy Model Design and Implementation

A multi-resolution model of visual attention have been proposed and implemented [5]. However, we have chosen to develop a region-based model due to a number of reasons. Firstly, eye fixations are directed towards regions of the image being viewed [6]. Secondly, region-based processing may utilise previous region-based psychophysical experimentation [1].

Wolfe [1] and Senders [11] list some of the region features so far discovered to affect the visual search path of a viewer:

- luminance contrast at the boundary of a region;
- hue contrast at the boundaries of different hued regions;
- contour concentration differences can lead to preattentive texture segmentation, influencing visual search [9];
- size differences between regions;
- motion is considered one of the strongest attractors of attention;
- depth cues, especially foreground/background effects.

Work has been carried out into the relationships of some of these features [6], but as yet there is no mention of a model that completely describes these features, their ordering and their relative weights. Our aim is to construct a more complete fuzzy preattentive feature relationship model by allowing for global effects and the use of contour concentrations to include some texture factors. The attention model fuzzifies the mean feature differences around segmented regions into three functions: Low, Medium and High.

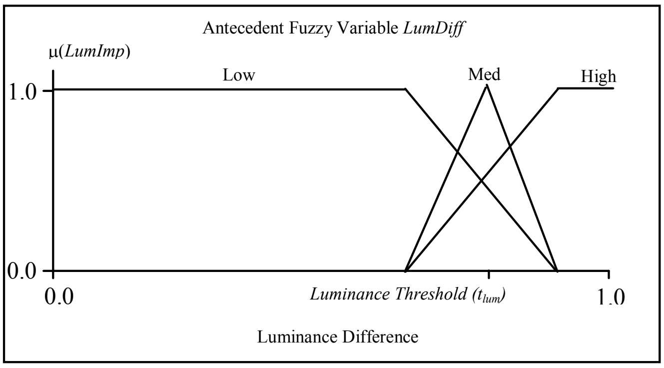

Fig. 2. An example of adaptive membership function approach, with the three membership functions centred around the mean luminance differences (t_{lum}).

Adaptive membership functions have been developed to model global feature influences. This has been implemented by passing the mean of the background feature differences t_{feat} to the system as a membership function shape parameter, so that the fuzzy threshold is dependent on the background variation in the image. This can be illustrated by two extreme cases. Scenes with low values of background activity within the image have the threshold as being the *just noticeable difference* values ascertained for simple single stimuli (around one percent contrast for a grey level edge on a constant background [12]). The other extreme is the case of a highly variant background, for example, a checkerboard. In this case, even the highest possible local contrast will not allow the region to become conspicuous, and so the mean difference value t_{feat} forces the threshold to the far right of the domain.

This threshold function models the sigmoidal effects analysed by Nothdurft [8], with regards to global saliency. His results show that the position on the feature domain varies according to the differences present in the background distractors. However, we have differed by considering the model to be sigmoidal in nature, unlike other systems [5, 6]. An example adaptive threshold function for the luminance features is shown in Fig. 2.

Trapezoidal membership function shapes have been used in our implementation for the sake of efficiency, due to their modelling previous experimental results [8] and the insensitivity of fuzzy systems to membership function shape, as compared to membership function position [13]. Other experiments we performed indicated a small area of uncertainty around the threshold [2], and so we have added the third medium term to membership functions.

These general threshold design principles have been implemented in a number of the membership functions for features that rely upon differences to attract the attention of the viewer: hue, size, and contour concentration. We present these in a summary form, due to their similarity to the luminance feature membership function.

Hue difference is modelled on the hue angle difference of the region and its surrounds. This is based on the assumption of the region being visually salient due to it being a different hue category to its surroundings.

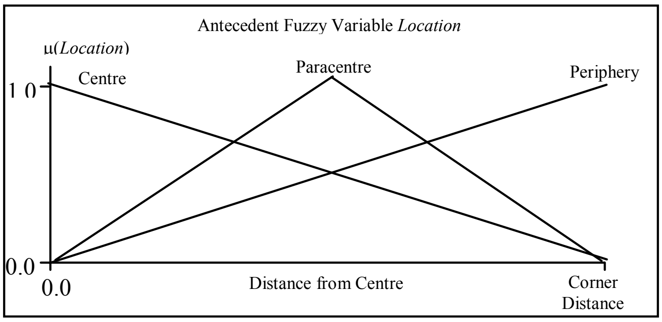

Fig. 3. Example nonthreshold Location feature membership function.

Size difference is modelled by the difference between the ratio of the image taken by the region, and the average ratio taken by its surroundings.

Contour concentration is modelled in a similar manner, using the DCM information for the region. A contour analysis approach called the *Directional Coherence Map* (DCM) [10] has been utilised to ascertain the contour strength, number and curvature estimates for the fuzzy logic importance system to use in its calculations.

Other features, such as location and background/foreground, have been found to influence viewer eye movements. These features do not require a threshold parameter to allow for global effects, as they are inherently absolute in nature. Fig. 3 illustrates the general design of these functions with three terms fuzzified over the domain of the variable.

The location feature is the spatial distance from the centroid of the image, with fuzzified terms Centre, ParaCentre and Periphery. Foreground/background segregation is modelled using the proportion of region segments that are on the border of an image. If the proportion is High, then the region will be considered to be a part of the background of the image.

These fuzzy membership functions are then used in the implication process to ascertain the visual importance of regions.

Region Model Implication Methodology

The following is an example list of some rules that constitute the fuzzy inference system for the region-based visual importance:

IF *LumDiff* IS High THEN *FinImp* IS High
IF *RegLoc* IS Centre THEN *FinImp* IS High
IF *EdgeProp* IS Low THEN *FinImp* IS High

The medium and low rules are similar to the high rules for the final importance variable *FinImp*, and are shown in a summary form:

IF *LumDiff* IS Med THEN *FinImp* IS Med
IF *LumDiff* IS Low THEN *FinImp* IS Low

The implication method uses a min operator. A bounded sum multiple aggregation method is used [13]. The use of multiple aggregation enables the modelling of the importance as being an additive contribution of activation in a number of feature dimensions. Along with the multiple aggregation method, a set of weights has been implemented for each of the rules. For now these weights are equal, except for the foreground/background feature, which has been made 2.5 times the others, to enhance foreground/background differentiation. Finally, the weighted fuzzy mean defuzzification method is used to obtain a crisp importance value.

The region importance is stored in a region importance map, as indicated in Fig. 1 and Fig. 4. Values inside the region importance map are normalised to [0.0, 1.0] for use by the adaptive rendering system, in order to bias its sampling regime.

IV. Integration of Importance Models into an Adaptive Renderer

The process of image synthesis can be generalised as a sampling problem. These samples take the form of pixels. Essentially, the image synthesis problem is one of making enough samples to correctly colour a pixel so that it represents a synthetic scene to the highest level of accuracy. One method of rendering is *ray-tracing*. This method simulates the process of light rays hitting the eye, in reverse. A ray (vector) is cast from a viewing position, through the location of a pixel in an image, into the synthetic scene. If this ray intersects a visible object, then the pixel the ray has passed through is coloured according to a formula linking the material properties of the object and a synthetic lighting model for the scene. Complex scenes incur a high computational cost when taking these samples, and so a major area of research in image synthesis is sample reduction. From this we derive two major points that need to be traded off in image synthesis, efficiency and fidelity. We present our work as a method of facilitating image synthesis efficiency by giving away some fidelity. This fidelity loss is ameliorated by the use of visual importance mapping to control where the degradation will occur.

We have incorporated the described visual importance model into a ray-tracing system. The sample generator stores its results in a quadtree, representing how spatially refined the image is at any point in time. The fuzzy importance system refers to this quadtree to ascertain the visual importance of a segment of the image, thus controlling the number of samples generated in that region. The segmentation is refined until the quadtree is refined to the level of a pixel. At this point, the pixels are super-sampled in order to produce the final high quality image.

The region-based visual attention model has been implemented in the ray-tracing system, to control the super-sampling of pixels. Super-sampling occurs by subdividing a pixel into segments and firing a ray at each vertex of the subdivision. This method removes some of the aliasing effects produced by discretely sampling the geometric model comprising the scene. We have implemented a perceptual super-sampling scheme for comparison, with or without a region importance bias. The perceptual super-sampling method used is one developed by Neumann[14]. We show in Fig. 4 a comparison of images generated using this method, and also the relative difference images between a work image and the importance biased images.

As well, we have a table showing the number of samples needed to render the images. Times are shown relative to the perceptual work image execution time. We also quantify the error between the work images and the degraded images using L_1 and L_2 norm ratios derived from the norm of the error image divided by the norm of the work image[10].

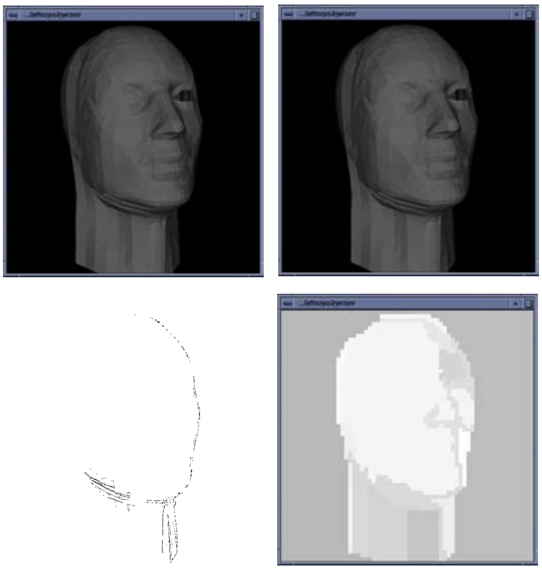

Fig. 4. A series of head images comparing the two methods of super-sampling. The work image is top left the biased image top right. The bottom row shows the difference between the two on the left and the right image illustrates the results of region importance calculations. More important regions are shown in lighter shades of grey.

TABLE I

RESULTS FOR UNBIASED AND BIASED PERCEPTUAL SUPER-SAMPLING RESULTS.

Method	**Samples Taken**	**Average Samples/Pixel**	**Relative Time**	**L1 ratio (L2 ratio)**
Perceptual Work Image	3,639,552	13.83	1.00	-
Perceptual Region Biased	2,642,688	10.04	0.62	0.004577 (0.009601)

V. CONCLUSION

In this paper we have described a novel fuzzy logic model of region-based visual attention. We have discussed its threshold approach and inclusion of contour concentrations as a measure of importance. We have then shown some examples of its application to image synthesis using ray-tracing, highlighting the savings in samples and the minimisation of the perceptual image distortion.

This adaptive rendering system is expected to be useful in applications in entertainment, where fidelity may be compromised for efficiency if the overall impression of the scene is not compromised.

Further work may be carried out in a number of areas. The system could be improved by the inclusion of empirically derived weights from eye tracking experiments, for the different feature dimensions. The membership functions could be modified to incorporate more low-level perceptual factors, such as contrast sensitivity and the use of perceptually uniform colour spaces. We also plan to extend this model to incorporate motion importance, and apply it to relevant aspects of computer animation.

REFERENCES

[1] J. Wolfe, "Visual Search: A Review," in *Attention*, H. Pashler, Ed. University College London Press, London, UK, 1996, pp. 13-73.

[2] R. Brown, B. Pham, E. Aidman, and A. Maeder. "Efficient Image Rendering Using a Fuzzy Logic Model of Visual Attention," *Proceedings of Advances in Intelligent Systems: Theory and Applications (AISTA)*, IOS Press, Amsterdam, Holland, pp. 314-319, Jan. 2000.

[3] J. Painter and K. Sloan, "Antialiased Ray Tracing by Adaptive Progressive Refinement," *Computer Graphics*, vol. 23, pp. 281-288, July 1989.

[4] H. Yee, S. Pattanaik, and D. Greenberg, "Spatiotemporal Sensitivity and Visual Attention for Efficient Rendering of Dynamic Environments," *ACM Transactions on Graphics*, vol. 20, pp. 47-54, Jan. 2001.

[5] L. Itti and C. Koch. "A Comparison of Feature Combination Strategies for Saliency-Based Visual Attention Systems," *Proceedings of Human Vision and Electronic Imaging IV*, SPIE - The International Society for Optical Engineering, Bellingham, USA, pp. 473-482, Jan 1999.

[6] W. Osberger and A. Rohaly. "Automatic Detection of Regions of Interest in Complex Video Sequences," *Proceedings of Human Vision and Electronic Imaging VI*, SPIE - The International Society for Optical Engineering, Bellingham, USA, pp. (To Be Published), Jan 2001.

[7] R. Hayasaka, J. Zhao, and Y. Matsushita. "Outstanding Objects-Oriented Color Image Segmentation Using Fuzzy Logic," *Multimedia Storage and Archiving Systems II*, SPIE - The International Society for Optical Engineering, Bellingham, USA, pp. 303-314, Nov. 1997.

[8] H. Nothdurft, "The Role of Features in Preattentive Vision: Comparison of Orientation, Motion and Color Cues," *Vision Research*, vol. 33, pp. 1937-1958, Sep. 1993.

[9] B. Julesz and J. Bergen, "Textons, the Fundamental Elements in Preattentive Vision and Perception of Textures," *The Bell System Technical Journal*, vol. 62, pp. 1619-1645, Jul. 1983.

[10] B. Guo. "Progressive Radiance Evaluation Using Directional Coherence Maps," *Proceedings of SIGGRAPH*, Addison Wesley, New York, pp. 255-266, Jul. 1998.

[11] J. Senders. "Distribution of Visual Attention in Static and Dynamic Displays," *Proceedings of Human Vision and Electronic Imaging II*, SPIE - The International Society for Optical Engineering, Bellingham, USA, pp. 186-194, Feb. 1997.

[12] B. Wandell, *Foundations of Human Vision,* Sunderland, USA: Sinauer, 1995.

[13] R. Berkan and S. Trubatch, *Fuzzy Systems Design Principles, Building Fuzzy IF-THEN Rule Bases,* New York, U.S.A.: IEEE Press, 1997.

[14] L. Neumann, K. Matkovic, and W. Purgathofer, "Perception Based Color Image Difference," *Computer Graphics Forum*, vol. 17, pp. 233-241, Sep. 1998.