Efficient Image Rendering Using a Fuzzy Logic Model of Visual Attention

Ross Brown[a], Binh Pham[a], Eugene Aidman[b] and Anthony Maeder[c]

[a]*Faculty of Information Technology, Queensland University of Technology, Australia.*
[b]*School of Behavioural and Social Sciences and Humanities, University of Ballarat, Australia.*
[c]*School of Electrical and Electronic Systems Engineering, Queensland University of Technology, Australia.*

Abstract

Physiological research has revealed constructs named “feature detectors” in the human visual system. These detectors react to differences in visual features, marking them as salient and thereby attracting the attention of the viewer. In this paper we outline a fuzzy logic system which processes 3D scene descriptions to compute the relative visual importance of regions in the scene. We detail experiments carried out to obtain parameters for the model, and the methodologies used in the aggregation, implication and defuzzification processes. We finish with a discussion of future work for further refinement of the model. The system is expected to have applications in the area of efficient image synthesis, in particular, progressive rendering and transmission techniques.

1. Introduction

Physiological research has discovered evidence for primate vision system constructs, called “feature detectors”, that are particularly sensitive to the orientation and movement of edges [7]. These and other similar detectors in the Human Visual System (HVS) are believed to interact in the early “preattentive” stage, causing regions of the viewing field to be identified as perceptually more salient, thus attracting the attention of a viewer [17].

In a previous paper [3] 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 complete hierarchy and integration models for preattentive visual features [17]. Preattentive feature-based models have been applied successfully to image processing, in particular, to the areas of active vision systems [8, 12], progressive transmission [18] and image compression [6, 14]. The active vision systems process images attempting to model the movement of the human eye. This attention modelling is performed by deducing the importance of regions based upon differences in the region features, integrating them together into an importance map to indicate the relative visual saliency of that region. Progressive image transmission systems have been developed [18], facilitating image discernment with a limited perceptual bias based upon contrast values, the higher contrasting areas being sent first.

More relevant to our work is the area of image compression, as this often applies a visual importance model to some form of region segmentation using either a crisp computational model [14] or a fuzzy system [6,11]. The use of the fuzzy system developed by Marichal et. al. [11] is a simple feature model limited to low bandwidth video systems. By tracking the size and motion of selected regions, the system is able to vary the compression rates in a video stream to maintain an optimum frame rate. With the system developed by Hayasaka et. al. [6] the segmentation is performed on still images to bias the JPEG compression algorithm to heavily compress the less important regions of the image. Their scheme is essentially data driven, with membership functions being derived from histogram analysis of the luminance channel of the image. The simple feature integration scheme is efficient, consisting of a weighted And of both fuzzified and crisp feature values derived from the image. Our system differs from the previous fuzzy systems by having an entirely fuzzy rule base, with no use of crisp values except for modification of membership function shape. Secondly, we seek to have a more complete model of feature integration based upon psychophysical information. Lastly, we novelly apply this fuzzy feature model to the area of 3D scene description decomposition, for the purpose of facilitating image synthesis efficiency.

Present adaptive image synthesis systems [5, 10, 2] do not decompose scene descriptions by the application of a preattentive feature model. These systems use image-space attributes to adaptively change the number of rays cast per pixel, using volume segmentation schemes [9]. Bolin and Meyer's [2] work involves the introduction of a model of early human vision to exploit visual masking and inferior colour fidelity in human vision to guide the dissemination of rays in a ray tracing system, they report modest increases in rendering efficiency. However, no integrated models of preattentive features have been developed to determine the visual importance of image-space regions using object-space features.

A scene decomposition system containing a fuzzy visual feature model would, once developed, facilitate image synthesis efficiency by applying the highest level of detail to the most perceptually salient regions of the scene. This would aid quicker discrimination of scene composition by the use of a perceptually modified rendering algorithm such as progressive raytracing [10], or progressive transmission [18]. This visual feature model will be more complete than those presently developed due to a less arbitrary feature integration rule set, and should have applications in other areas needing a feature-based visual attention model.

Extending and applying the previous work into new areas, we describe in this paper the design, implementation and evaluation of such a decomposition system that parses a 3D scene description into regions. 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, including the utility of fuzzy systems in modelling human behaviour, and the experimentation behind the parameters used in the system. A detailed example is then shown of the development of a fuzzy rule base for the luminance feature. The paper then concludes with a description of methodologies to extend and integrate size and depth features into the system.

2. System Conceptual Framework

The decomposition system is constructed from three major components: a *scene description parser*, a *perceptual object geometry representation* and a *fuzzy logic visual importance model*.

The scene description parser decomposes a scene containing the geometry, using OpenGL™ to project the 3D geometry to the 2D viewing plane, where the visual importance calculations are performed. The viewing plane is segmented into quadrants, implemented as a modified quadtree [16]. The 2D geometry is merged into perceptual objects according to Gestalt grouping rules [15]. These rules allow the modelling of *perceptual grouping*, which occurs when spatially close objects have similar attributes. For example, a large X made up of small Xs is perceived at certain scales as being the larger grouped X, or as the smaller component Xs.

A fuzzy logic model then assesses the visual importance of the perceptual objects, integrating both local and global scene statistics and user supplied parameters. Local differences in preattentive visual features can cause a scene region to become perceptually salient [17], for example, due to a sharp luminance contrast. This salience effect is modelled by the mean feature value difference between a perceptual object and its local neighbourhood. In addition, it has been shown that global scene feature variability enhances or suppresses the previously mentioned saliency [13], that is, the salience of the target is influenced by the surrounding visual noise. The global effects are modelled using the standard deviation of the features across the entire scene. We use fuzzy membership functions [1] due to their ability to model human reasoning. The perceptual saliency of an area can only be coarsely evaluated by humans, therefore the system fuzzifies the feature difference and standard deviation variables, using the linguistic terms High, Intermediate and Low. The parameters and characteristics of these membership functions are drawn from our own psychophysical experimentation with luminance, size and depth features and from relevant psychological literature.

3. Fuzzy System Design and Implementation

We have sought to model these perceptual effects by fuzzifying the mean feature differences around the perceptual object in question. Membership function parameters for this variable are drawn from psychophysical experimentation that elicits a subjective assessment of region importance produced by differences in luminance, size or depth features. Due to the sigmoidal nature of perceptual salience [13], the subject is asked to evaluate the conspicuousness of a target region using the linguistic terms High and Low.

A frequency histogram from the experiment is used to characterise these linguistic terms, as they relate to the magnitude of local differences in features. An Intermediate term is statistically derived from the threshold region of confusion between the High and Low terms. We describe this process in detail for the luminance feature.

3.1 Luminance Feature Example

A preliminary experiment was performed with 4 subjects situated in a standardised room, with a calibrated monitor [4]. The subject was shown a grid of 9 quadrilaterals for 100 milliseconds, followed by a visual noise mask for one second. The subject was asked to rate the conspicuousness of the centre target (Figure 1) as High or Low by hitting, during the presentation of the noise screen, one of the two shift keys on a computer keyboard. Four sessions of 125 stimuli were shown with local target-surround differences of 0.0, 0.25, 0.50.75 and 1.0, with 0.0 representing no signal (black) and 1.0 representing the highest possible monitor signal (white).

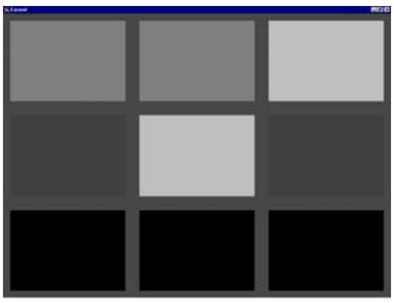

Figure 1 Example luminance experiment stimulus.

The two High, Low categories were chosen for a number of reasons. Firstly, we were seeking to gain an instinctive evaluation from the subjects of the conspicuousness of the target, therefore it would be difficult for the subjects to accurately assess anything more than a two level scale. Secondly, the phenomenon of visual popout is sigmoidal in nature with a threshold and steep change gradient, leading to a natural two level evaluation scale.

These experiments produced a frequency histogram relating the terms High and Low conspicuousness with the local mean difference in luminance. A linear membership function was then fitted to the top of the frequency data. Using a scatter plot, the distance from the spacing between the points indicates the level of agreement between the subjects in the study. The region where the values digress from each other the most is the region we label with the Intermediate fuzzy term (Figure 2). In addition, the subjects acknowledged making errors while judging the conspicuousness of the target. This added a constant noise value to the data that was removed, producing the fuzzy High, Intermediate and Low membership functions shown in Figure 3.

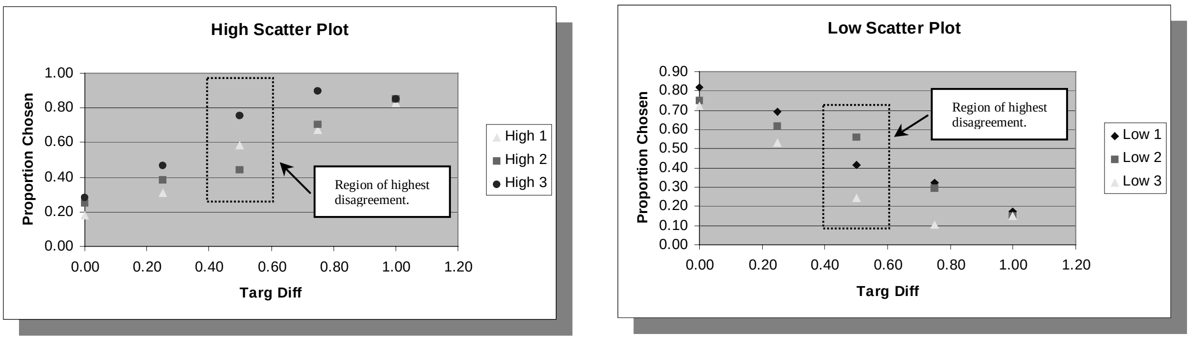

Figure 2 Scatter plot of proportion of highs and lows chosen for target differences, by each subject.

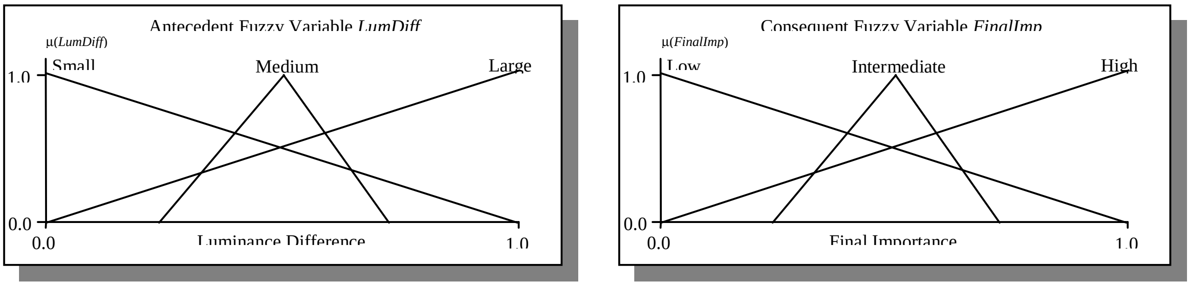

Figure 3 Membership functions for fuzzy variables LumDiff and FinalImp.

Furthermore, preliminary experimentation has shown that a variation in the standard deviation of the background luminances affects the nature of the membership functions. For example, a high standard deviation in the background suppresses the High membership function, while a low standard deviation in the background enhances the High membership function. These effects have been modelled using a combination of adaptive changes to the membership functions and the use of a varying weight on the High membership function.

3.2 Fuzzy Deduction Methodology

In the design of the system, consideration has been applied to the three components of fuzzy deduction: *aggregation*, *defuzzification* and *implication*. It has been observed that the interaction of preattentive features is competitive in nature [8]. Furthermore, objects that are unique in appearance due to a single feature are more conspicuous than objects that are unique in appearance due to a conjunction of features [17]. For example, a red circle among otherwise identical blue circles is more conspicuous than a red circle among red squares and blue circles.

We hypothesise a two-phase hierarchical rule base should be developed from these observations and our own visual experiments with combinations of preattentive features. The first level deals with the high importance rules, linking to the hypothesis about the competitive nature of preattentive features, that is, an object becomes strongly conspicuous if differing by only one feature. Essentially, if one of the feature dimensions is high in difference, then the object is high in importance, producing an ored rule combination:

IF *LumDiff* IS Large OR *SizeDiff* IS Large THEN *FinalImp* IS High

If the above rule does not have a degree of fulfilment over an alpha threshold, then a second set of rules is used, linked to the second part of the hypothesis. That is, if the region is not highly conspicuous due to one feature, then the conspicuousness of the region is low to intermediate in value, and may be modelled by an averaging of the features. The example below shows combinations of two features, size and luminance:

IF *LumDiff* IS Medium AND *SizeDiff* IS Medium THEN *FinalImp* IS Intermediate
IF *LumDiff* IS Medium AND *SizeDiff* IS Small THEN *FinalImp* IS Low
IF *LumDiff* IS Small AND *SizeDiff* IS Medium THEN *FinalImp* IS Low
IF *LumDiff* IS Small AND *SizeDiff* IS Small THEN *FinalImp* IS Low

An aggregation method has been chosen to reflect these rules. For the High rules a Minimum aggregation method is used, along with First of Maxima defuzzification. This defuzzification method was chosen due to its ability to mimic competitive systems, a feature of the preattentive phase of the HVS. The first of maxima is also highly sensitive to the adaptive rules that are used in this system.

For the second set of rules, a Maximum aggregation method is used along with the centre of gravity defuzzification method, which should emulate the averaging which occurs across the low-level importance rules in the system.

The implication method is derived from the results of the experiments with feature combinations. The simple example of the luminance feature indicates that the antecedent and consequent functions are essentially the same in shape (Figure 4). It would be nearly impossible for a subject to separate the concepts of feature differences from visual conspicuousness. It is expected that the relationship between the consequent and antecedent function will only change when other features are incorporated into the model. With multiple features we believe that a hyper-dimensional fuzzy surface will best reflect the feature interactions. Once the general nature of pair-wise interactions has been established, extra features may be added to the combinations by extrapolating the results of the first set of experiments. Extrapolation will be performed by analysing the data for consistent trends from combinations of the features. If the first hypothesis about the combinations of features holds, that is, the feature interactions are competitive for High rules and averaged for those regions without High levels of salience, then the implication method will be able to be generalised to incorporate other features.

4. Discussion

To conclude, we have found that our hypothesis about the nature of the behavioural response to local luminance contrast is supported by our data. Secondly, preliminary experiments have indicated support for the hypothesis of an effect induced by the variability of surrounding distracters, further confirming the previous psychophysical work in visual attention [13]. From these results we have constructed a fuzzy system to model this behaviour.

The intended accuracy of the system will be an influence on future work, due to the fuzzy logic model used in the system. The categories are innately coarse due to the sigmoidal nature of the phenomena to be modelled. Therefore, the focus of experimentation will be to accurately find threshold levels for each of the features and to discern the influence of background feature variability upon the nature of this threshold point. A major component of future experimental work will be concentrated on refining the models of threshold effects and the intermediate areas of subjective confusion. At present, the luminance feature has been characterised by preliminary experimentation. This luminance experimentation will be continued in a similar fashion to that described in section 3.1, and will extend to incorporating size and depth features.

Presently, the scene parser and quadtree segmentation modules have been implemented. Further implementation work will be carried out to complete the fuzzy importance model. We are endeavouring to implement two forms of system, one with 2D membership functions as shown in this paper, and another with 3D membership function surfaces. An evaluation will be performed to ascertain the utility of each system.

A major question is the matching of the subjective values of conspicuousness with the eye attracting ability of a region. The visual importance system implementation will be evaluated by calculating the statistical correlation between the locations of computed visual importance levels and recordings of human eye fixations. The test scenes used in these experiments have been simplified to remove higher level factors of object recognition, which could influence eye movements. We believe that this simplification will not affect the evaluation of the bottom-up stimuli driven model we have implemented.

Apart from image synthesis efficiency, other possible applications of the scene decomposition system are in the areas of scientific visualisation, vision simulation, entertainment and progressive data transmission in medical imaging.

References

1. R. Berkan and S. Trubatch, *Fuzzy Systems Design Principles, Building Fuzzy IF-THEN Rule Bases*, IEEE Press, New York, U.S.A., 1997.
2. M. Bolin and G. Meyer, Visual Difference Metric for Realistic Image Synthesis, *Proceedings of SPIE*, **3644**., Jan, San Jose, USA, pp. 106-120, 1999.
3. R. Brown, B. Pham and A. Maeder, A Fuzzy Model for Scene Decomposition based on Preattentive Visual Features, *Proceedings of SPIE*, **3644**, Jan, San Jose, USA, pp. 461-472, 1999.
4. CCIR, *Methodology for the subjective Assessment of the Quality of Television Pictures*, Document ITU-R BT.500-6, 1994.
5. N. Guivens, Adaptive Ray Casting (ARC) Algorithm for Simulation of Machine Vision Systems, *Proceedings of SPIE*, **1822**, pp. 409-418, 1992.
6. R. Hayasaka, J. Zhao and Y. Matsushita, Outstanding Objects-Oriented Color Image Segmentation Using Fuzzy Logic, *Proceedings of SPIE*, **3229**, Jan, San Jose, USA, pp. 303-314, 1997.
7. D. Hubel, *Eye, Brain and Vision*, Scientific American Library, New York, USA, 1995.
8. L. Itti and C. Koch, A Comparison of Feature Combination Strategies for Saliency-Based Visual Attention Systems, *Proceedings of SPIE*, Jan, San Jose, **3644**, pp.473-482, 1999.
9. K. Klimaszewski and T. Sederberg, Faster Ray Tracing Using Adaptive Grids, *IEEE Computer Graphics and Applications*, **17,** pp. 42-51, 1997.
10. J. Maillot, L. Carraro and B. Peroche, Progressive Ray Tracing, *Proceedings of 3rd Eurographics Workshop on Rendering*, May, Bristol, UK, pp. 9-19, 1992,
11. X. Marichal, T. Delmot, C. De Vleeschouwer, V. Warscotte and B. Macq, Automatic Detection of Interest Areas of an Image or of a Sequence of Images, *IEEE International Conference on Image Processing*, Sept., Lausanne, Switz., pp. 371-374, 1996.
12. R. Milanese, T. Pun and H. Wechsler, A Non-Linear Integration Process for the Selection of Visual Information, *In: Intelligent Perceptual Systems, Lecture Notes in Artificial Intelligence*, V. Roberto ed., 1993.
13. H. Nothdurft, The Role of Features in Preattentive Vision: Comparison of Orientation, Motion and Color Cues, *Vision Research*, **33**, pp. 1937-1958, 1993,
14. W. Osberger and A. Maeder, Automatic Identification of Perceptually Important Regions in an Image using a Model of the Human Visual System, *Proceedings of ICPR*, Aug., Brisbane, Australia, 1998.
15. I. Rock and S. Palmer, The Legacy of Gestalt Psychology, *Scientific American*, **263**, pp. 48-61, 1990.
16. H. Samet, *The Design and Analysis of Spatial Data Structures*, Addison Wesley, Reading, USA, 1990.
17. J. Wolfe, Visual Search: A Review, *In: H. Pashler ed., Attention*, University College Press, London, UK, 1996.
18. H. Zabrodsky and S. Peleg, Attentive Transmission, *Journal of Visual Communications and Image Representation*, **1**, pp. 189-198, 1990.