

Dual-Modality Two-Stage Direct Volume Rendering

M. G. Jones

Complex Active Visualisation Laboratory
School of Information Technology
Griffith University
Gold Coast, Australia
Michael.Jones@gu.edu.au

S. G. Nikolov

Centre for Communications Research
University of Bristol
UK
Stavri.Nikolov@bristol.ac.uk

Abstract – *A modified direct volume rendering algorithm is presented supporting simultaneous display of two source modalities via a spatially-variant Region-of-Interest function that weights the combination of individually-specified opacity transfer functions. The algorithm is applied within a focus + context paradigm. Examples are presented using computed tomography and colour histological image data from the Visible Human Project. The algorithm is implemented as a two stage process. During the first, computationally expensive, stage global raycasting generates an array of partial projection segments for each displayed pixel; during the second, image generation supplements segment combination with local raycasting around the ROI - this can be much more rapid. The visual data fusion approach is compared with conventional rendering and shown to provide an effective method of exploring complex 3-D multi-modality data.*

Keywords: volume rendering, focus + context, multi-modality, display, visualisation

1 Introduction

Focus + context visualisation techniques are widely used to scrutinise large and complex data sets [1, 2]. These techniques simultaneously provide an enhanced view within the *focus*, or region-of-interest (ROI), and a broader, non-enhanced view of the *context*, or surrounding region; the boundary between the two regions may be either abrupt or smoothly blended. Moving the focus position through a complex data set is one effective method of visual data mining [3]. The presence of the context allows the user to maintain orientation within a, potentially, very large data set.

A sub-class of focus + context techniques consists of those that change the displayed data dimensions within the focus. In this paper our first theme is consideration of dual-modality focus + context visualisation via direct volume rendering (DVR). For this, the focus region is provided,

principally, by one modality and the context region, by another. Via this route, data from one modality may be explored in the spatial context of another.

This rendering method become useful when applied interactively, with the position of the ROI controlled by the user (or, alternatively, controlled by some form of path-generation algorithm) so as to facilitate visual exploration of a data set. However DVR algorithms are computationally expensive. Our second theme is thus analysis of the decomposition of the rendering process into two stages: firstly, before display begins, a *base* component is generated from a non-enhanced view of the data (the context); secondly, during and throughout image display, a *supplemental* component is continually updated as the user moves the ROI. Since the ROI may occupy only a small fraction of the entire data volume, then very significant increases in rendering speed are achievable.

The methods presented are based upon a gaze-tracked, region-enhancement, volumetric rendering system constructed by the authors and previously reported [4, 5, 6]. We develop the DVR algorithms in section 2. Section 3 considers two-stage rendering. In section 4 the techniques are applied to render voxel data from the Visible Human Project (VHP) [7], combining colour histological data with grey-scale computed tomography (CT) data. Section 5 presents our conclusions. For demonstration of the technique, our method is compared with conventional single modality and dual modality rendering methods across a number of modality combinations.

2 Conventional and dual-modality DVR algorithms

A huge range of methods exists for rendering volumetric data in a 3-D fashion. A large subset of these could be used as a basis for multi-modality volumetric rendering, including surface fitting (SF) and direct volume rendering (DVR) approaches [8, 9, 10]. We use a front-to-back

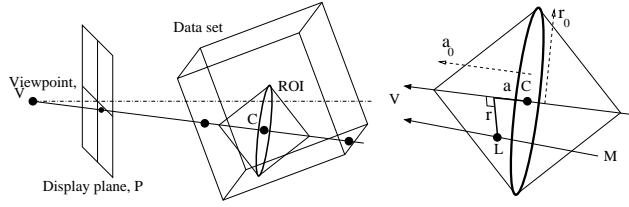


Figure 1: Projection geometry for F-to-B composition and ROI positioning and orientation within the data volume

(F-to-B) compositing method of DVR as the basis for our development. F-to-B compositing software and hardware implementations are certainly not the fastest methods, but they are simple, well characterised, and, operating in image space, have low levels of (well-studied) artifacts [11]. Presenting our approach with reference to this basic implementation should thus more clearly illustrate its characteristics. Our general approach of spatially-varied, multi-modality display could in future be applied to many other rendering algorithms, operating in image space or object space, such as splatting [12] or 3-D texture mapping [13].

Figure 1 shows our projection geometry. F-to-B composition can be considered a ray-tracing operation. For each pixel on the display plane, P , a ray is projected from the viewpoint, V (the eye position) through the pixel and on through the data set. The data set is multiply sampled along the ray at a set of I [usually] regularly-spaced positions, $i \in \{0..I-1\}$ and these data samples are mapped to [partial-]opacity samples, α_i , using a suitable transfer function, and to colour samples, c_i . These colour samples are generally calculated using the data colour, together with the directions and intensities of single or multiple light sources. When surfaces are rendered opaquely, reflections (diffuse and specular) and shadowing can provide great realism and clarify 3-D structure. However, when many, only partially opaque, samples are composited onto a single display pixel, much of the realism of complex lighting is lost. Whilst we concede that some complex lighting techniques are beneficial, such as the spectral volume rendering presented in [14], within this work, we currently use only ambient lighting calculations.

Compositing reduces the α_i and c_i samples to a final pixel colour. The formula for conventional (single modality) front-to-back composition can be written iteratively:

$$\begin{aligned} C' &= C + c_i \alpha_i T \\ T' &= T (1 - \alpha_i) \end{aligned} \quad (1)$$

where C is replaced by C' (and T by T') after each of I iterations.

A region-enhancement (RE) form of this can be developed in which a spatially-variant weighting function $\omega_{r,a}$ is used to blend between two opacity functions. Using the

shorthand notation, $\overline{\omega_{r,a}} = 1 - \omega_{r,a}$:

$$\begin{aligned} C' &= C + c_i [\overline{\omega_{r,a}} \alpha_i + \omega_{r,a} \beta_i] T \\ T' &= T (1 - [\overline{\omega_{r,a}} \alpha_i + \omega_{r,a} \beta_i]) \end{aligned} \quad (2)$$

Here, $\omega_{r,a}$ is a function, with a range of $[0, 1]$, which peaks at C and decreases with distance from C along the radial (r) and axial (a) directions. With reference to Figure 1, $\omega_{r,a}$ is non-zero only within the ROI, which is shown with the form of a double-cone, oriented towards V . The size of this double-cone, in radial and axial dimensions, is given by parameters r_0 and a_0 , respectively. There are now two opacity transfer functions: one which produces α_i ; the other which produces β_i . Varying these transfer functions affects the relative appearance of the focus and context regions; varying the form of $\omega_{r,a}$ affects how these two are blended together and the size of the ROI. We currently use a simple weighting function, where $\omega_{r,a}$ is the value, clamped to a range of $[0, 1]$, of:

$$1 - \frac{r}{r_0} - \frac{a}{a_0} \quad (3)$$

When two modalities are present, a straightforward extension becomes possible, by considering the second transfer function to be applied to the second modality. Thus:

$$\begin{aligned} C' &= C + [c_i \overline{\omega_{r,a}} \alpha_i + d_i \omega_{r,a} \beta_i] T \\ T' &= T (1 - [\overline{\omega_{r,a}} \alpha_i + \omega_{r,a} \beta_i]) \end{aligned} \quad (4)$$

Here, the second modality has direct analogs of c_i and α_i : these are d_i and β_i respectively. Similarly to the single modality region enhanced form, $\omega_{r,a}$ acts to blend between the two regions.

Since the range of ω is $[0, 1]$, this allows no contribution of the focus modality outside the boundary of the ROI and no contribution of the context modality at the centre of the ROI. A slight modification, allows for a degree of contribution to each to be added, via two global parameters, l and h :

$$\begin{aligned} C' &= C + [c_i \overline{\omega'_{r,a}} \alpha_i + d_i \omega'_{r,a} \beta_i] T \\ T' &= T (1 - [\overline{\omega'_{r,a}} \alpha_i + \omega'_{r,a} \beta_i]) \end{aligned} \quad (5)$$

where,

$$\omega'_{r,a} = l + (h - l) \omega_{r,a} \quad (6)$$

(and similarly to before, $\overline{\omega'_{r,a}} = 1 - \omega'_{r,a}$). The range of ω' is thus $[l, h]$, where $0 \leq l \leq h \leq 1$.

3 Two-stage rendering

Some focus + context algorithms can be decomposed into two stages, yielding a base component that is calculated throughout the entire data volume and a supplemental component that is calculated only within the ROI. As the ROI moves during interactive visualisation, only the supplemental component need be updated, providing the

viewpoint and data volume remain fixed. However, the use of motion, via, for example, rotating the data volume, is hugely beneficial: highly significant depth cues are provided by motion parallax and kinetic depth [15] and the spatial relationships between features become far clearer. Two-stage rendering may be applied with motion, if the viewpoint / data volume trajectory is fixed - rather than allowing the observer to arbitrarily spin and move the data volume, it is shown rotating back and forth over a set number of rendered frames. Base components would then be generated for each frame in this sequence and cycled through during display.

A most basic scenario might consist of the base component being a simple image without enhancement and the supplemental component being an image patch covering the projected outline of the ROI; blending might be used to fuse these to form the final image. In general such a simple decomposition is not possible, since simple image patches contain no information about relative depth. It may, for example, be necessary for the base projection to comprise volumetric, rather than 2-D, data.

The left of Figure 2 shows the situation for two-stage dual-modality DVR. Essentially the problem is one of substituting a replacement ROI segment into the base component calculated with no ROI. Each ray-traced line may be considered to have three segments: a leading segment (nearest the viewpoint); a ROI segment; and a trailing segment. We introduce the following definitions:

- C_L Leading segment partial colour contribution
- T_L Leading segment residual transmission
- C_R ROI segment partial colour contribution
- T_R ROI segment residual transmission
- C_T Trailing segment partial colour contribution

These may be considered as the final values of C and T if the equations of section 2 were iteratively applied over each segment individually (and in each case initialising with $T = 1$, i.e. the segment is fully exposed rather than partially hidden). The final pixel colour may then be calculated as:

$$C = C_L + T_L (C_R + T_R \cdot C_T) \quad (7)$$

The essential problem with equation 7 is that since the substituted segment may be positioned anywhere within the original segment, it is necessary to store values for C_L , T_L and C_T at every projected sample position. This may require many megabytes of storage. One of several alternative approaches is to store only every n^{th} sample; we term n the partial storage factor. This is represented on the right of Figure 2 for the case $n = 4$. The filled black circles within the ROI represent the samples which must be calculated in all cases. The filled grey circles represent the sample positions for which C_L , T_L and C_T are stored. The unfilled circles represent the remaining sample positions. It is thus necessary to calculate the ROI segment, beginning from the nearest available stored position, including any unfilled cir-

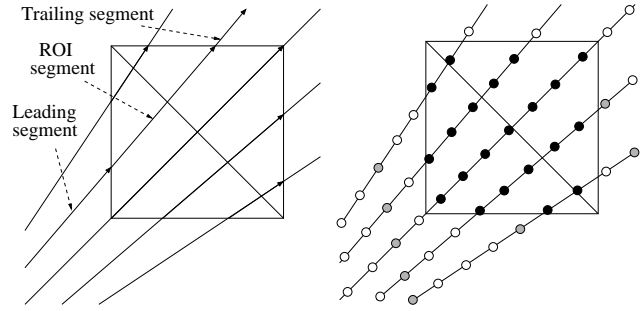


Figure 2: Two-stage rendering: (a) combination of leading segment, ROI segment and trailing segment; (b) sample positions and partial storage of base component calculations (see main text).

| n | $r_0 = 0.5$ $a_0 = 0.5$ | $r_0 = 0.25$ $a_0 = 0.25$ | $r_0 = 0.125$ $a_0 = 0.125$ |
|-----|----------------------------|------------------------------|--------------------------------|
| 1 | 23.7% | 2.96% | 0.370% |
| 4 | 24.2% | 3.10% | 0.404% |
| 16 | 26.4% | 3.65% | 0.543% |

Table 1: Percentages of the total samples required for calculation of the supplemental component as a function of ROI size and partial storage factor, n .

cle positions as necessary - essentially substituting a longer segment, a portion of which duplicates the original. This decreases the efficiency of the decomposition somewhat, but can greatly reduce the storage requirements. Table 1 shows typical percentages of the total ray-traced samples that must be calculated for different sizes of ROI, when using three different values of n . The entries in this table were calculated for a 384^2 pixel rendering of a 384^3 voxel data cube viewed directly face-on at a distance that results in the projected cube boundary just brushing the edge of the display window, with a horizontal [=vertical] field of view of 40 degrees, a projection sample spacing equal to the cube voxel spacing and the ROI positioned centrally in the data cube.

The table shows that for the ROI sizes we typically use (central column), calculation of the supplemental component requires perhaps only a few percent of the samples required without the two-stage approach. The relationship between these figures and algorithm time savings is heavily implementation dependent (and critically sensitive to issues of cache organisation and memory access patterns). However we generally anticipate at least an order of magnitude increase in rendering speed.

4 Examples and Applications

The example images we present here are drawn from the field of medical imaging. However there are many fields that also require scrutinisation of large multi-modality, or multi-spectral, data sets - for example in oil exploration or other tasks of geophysical mapping. We suggest that the common requirement making worthwhile an approach such as ours is that data interpretation should be dependent upon the spatial position at which certain measurements occur. A typical example from medical imaging is of combined functional and anatomical volume visualisation [16]. In functional magnetic resonance imaging (fMRI), knowing the position within the brain at which a heightened response occurs is critical in interpreting its relevance. The relative low-resolution of fMRI data does not lend itself to effective volumetric rendering, but if rendering were performed in conjunction with much higher resolution anatomical information - such as provided by conventional MRI data, a synergy may result. We have chosen to present hybrid renderings of colour histological and [frozen-cadaver] CT data from the Visible Human Project. These data represent high resolution spatial information, with at least some degree of complementarity between them.

4.1 Conventional and dual-modality DVR rendering

Figure 3 shows orthogonal cut-slice displays through the two modalities used for this work. The upper panes show the colour histological VHP data from the male head; the lower panes show the corresponding CT data. Manual adjustment of the relative displacement of the two source data sets was used to obtain approximate registration, but since the emphasis in this work is on the rendering process, more advanced and accurate transformation-optimisation techniques were not applied.

A reasonable starting point for evaluating our dual-modality DVR method is to compare conventional DVR rendering of the individual source modalities. Figure 4 shows perspective projection views through each of these.

An aspect of our dual-modality DVR algorithm (equation 4) is its similarity to region-enhanced DVR (equation 2). RE rendering of the two source data sets is shown in Figure 5. Within the figure, the ROI has been moved to three different positions within the data volume to illustrate the RE effect and its use in interactive data exploration.

The effect of our dual-modality algorithm is shown in Figure 6. The upper panes demonstrate the use of maximal blending between the histological and CT data (equation 4) - in this case none of the histological data is seen outside the ROI; the lower panes show the effects when the blending range is constrained (see equation 6) - in this case within a range of $[0.016, 0.5]$.

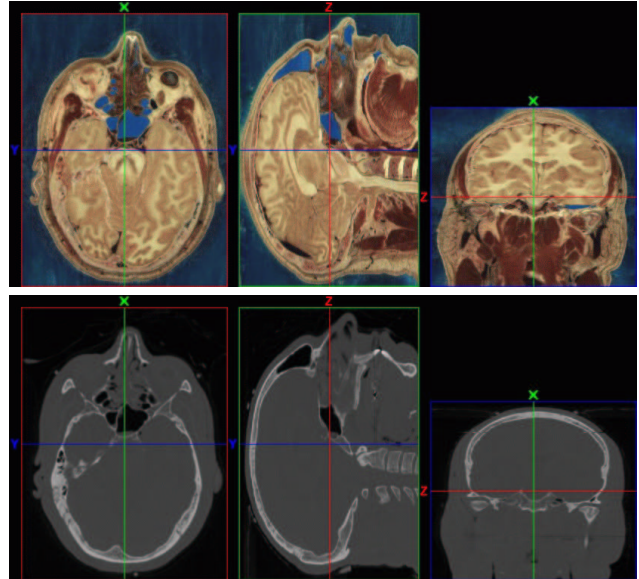


Figure 3: Orthogonal slice cuts through the two source data sets fed to the rendering engine. Upper panes show the colour histological image data; lower panes show the corresponding CT data.

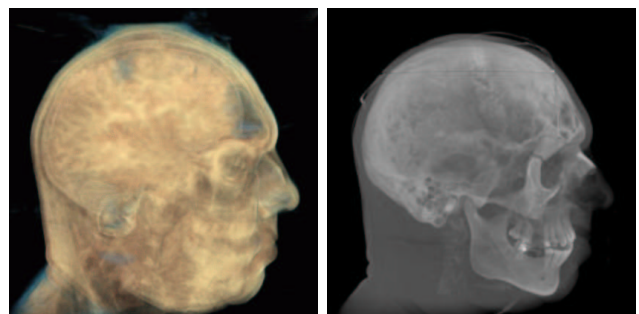


Figure 4: Conventional DVR of colour histological data and grey-scale CT data.

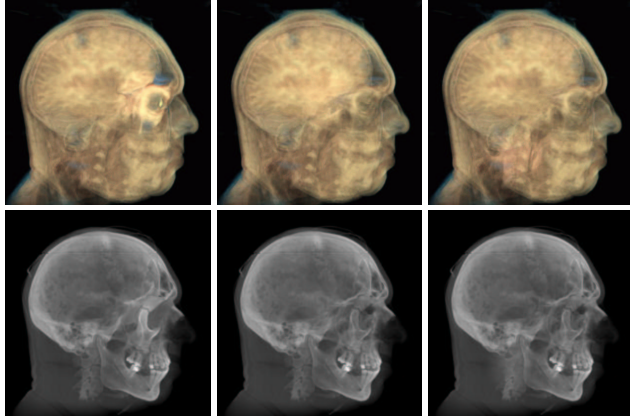


Figure 5: Region-enhanced DVR of colour histological data and grey-scale CT data for three ROI positions.

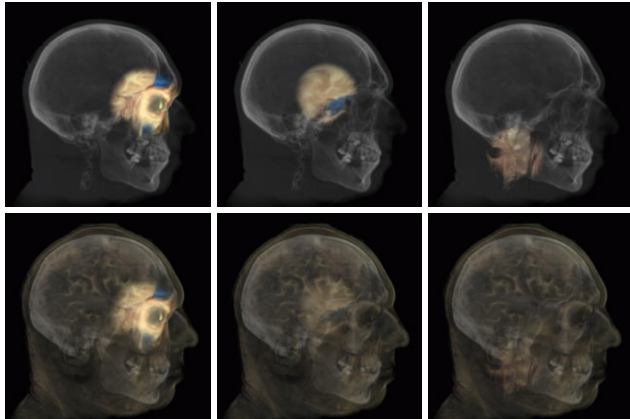


Figure 6: Dual-modality DVR of colour histological data (focus) and grey-scale CT data (context). Upper and lower panes show the effects of varying ω' over ranges of $[0, 1]$ and $[0.016, 0.5]$, respectively.

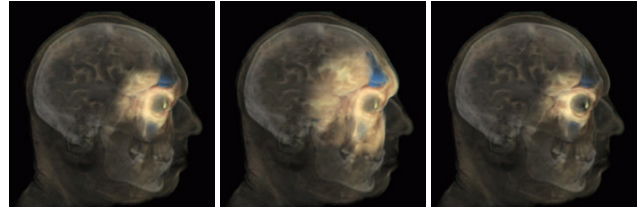


Figure 7: Effects of different sized regions of interest. Left pane: $a_0 = 0.15$, $r_0 = 0.25$. Middle pane $a_0 = 0.15$, $r_0 = 0.4$. Right pane $a_0 = 0.08$, $r_0 = 0.25$.

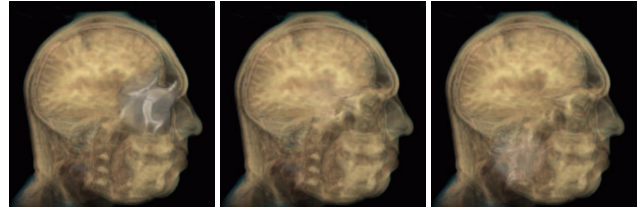


Figure 8: Dual-modality DVR of grey-scale CT data (focus) and colour histological data (context).

4.2 Variation in weighting function form and further multi-modality fusion

The size and shape of the weighting function may have a great effect on the rendered images. Figure 7 shows the effects of changing the extent of the ROI in the radial (r_0) and axial (a_0) directions. It is seen that increasing the radial extent (middle pane) produces a much wider focus. Decreasing the axial extent (right pane) produces sharper features in the focus since less overlapping features in the data volume contribute.

The context and focus modalities may be swapped and the effects of doing this are shown in Figure 8. Subjectively we feel that this display is not as useful. We suggest that the reason for this is that the context modality should be one that provides clear structural information rather than excessive detail.

The main purpose of the context view of the entire volume is to provide spatial reference. As such, the fidelity of this region need not be that high and a simplified version of the context might suffice. Indeed, a simplified version of the context might even be superior, since obscuring detail away from the ROI is removed. One way of achieving this is through a further modification of the rendering algorithm, to throw away low-level detail from the context modality. A somewhat crude approach to doing this is as follows.

The top line of equation 4 may be written:

$$C' = C + [c_i \overline{\omega_{r,a}} \alpha_i + d_i s_i] T \quad (8)$$

where, $s_i = \omega_{r,a} \beta_i$. Our context simplification approach

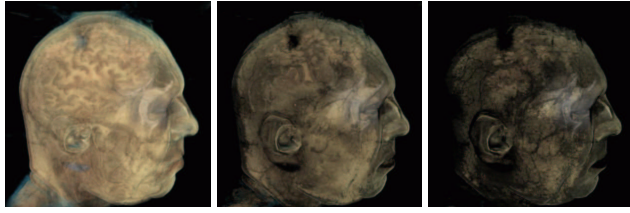


Figure 9: Effects of increasing (from left to right) the threshold s_t within dual-modality DVR applied to CT data (focus) and colour histological data (context).

is to compare s_i against a threshold s_t and, if it is less, assign s_i a value of zero. In this way only the sufficiently opaque features in the context modality are seen. The effects of this are shown in Figure 9.

The basic image in Figure 9 certainly shows significant artifacts owing to the abrupt removal of the more transparent features and does not provide an accurate representation of the context modality. It may be, however, that in keeping with the ideas of non-photorealistic rendering [17], this is an advantage when visually exploring complex data.

5 Conclusion

The technique presented allows one modality to be examined in the spatial context of another. The images certainly show elements of both modalities and the blend between them is smooth. A full and quantitative assessment of the benefits of the technique is still to be done, however. It is not clear how the human visual system responds to multi-modality images of this type and the degree to which the context information is noticed may be low, since when concentrating in one part of the image, the useful visual field size can be small[18]. The role of the surrounding context modality may thus be questioned. However the size of the useful visual field size is known to be very task-dependent - and so relevant structural information may still intrude into awareness in some instances. Also, a proportion of each modality may be maintained both with the centre of the ROI and the distant surround, so that a degree of the natural awareness of each modality, singularly, is kept.

Implementation of the method is on going. Whilst the software renderer fully supports all the functionality reported here, it does not have performance suitable for real-time, interactive volume navigation [19]. A base component projection with detail similar to shown in the figures (384^2 pixels) takes of the order of ten seconds using the current software implementation. Using the two-stage decomposition and re-writing the software to use the much faster calculation afforded by single-instruction multiple-data (SIMD) instruction sets, we hope to achieve real-time operation with limited resolution display in the near future. Further optimisation may be possible by exploiting the rel-

ative insensitivity of the human visual system to moving features; approximate, rather than strictly accurate rendering can be applied for these. An example of this approach of perception based rendering is given in [20].

The comparison between rendering of the pre-fused data and the dual-modality DVR blending shows how subtleties in the focus modality are preserved within the displayed image. Our conclusion is that the most useful application of the method would arise when a modality containing local subtlety but global commonality could be used as the focus modality, with the context modality being one that exhibits gross features and structure that provide spatial reference.

The future direction envisaged is to develop the rendering methods to take account of measured characteristics of human visual perception. The broad approach used to guide such development is common to that of non-photorealistic rendering; the conveyence of information is taken to be more important than the physical accuracy of the rendering model employed. Hence, if artificial representation of, say, structural information in the periphery of vision more effectively maintains the observer's sense of orientation, then that will be adopted.

Acknowledgements

Thanks are due to The Royal Society (UK), whose Paul Instrument Fund provided a project grant for construction of a stereoscopic gaze-tracked display, making possible much of the work presented herein, and which supported the second author via a conference grant. The first author also wishes to gratefully acknowledge the support provided within his previous position at the Department of Medical Physics and Bioengineering, United Bristol Healthcare NHS Trust, UK, where the majority of the research was carried out.

References

- [1] S. G. Nikolov, M. G. Jones, D. R. Bull, C. N. Canagarajah, M. Halliwell, and P. N. T. Wells, "Focus+context visualisation for image fusion," in *4th International Conference on Information Fusion (Fusion 2001), Montreal, Canada, 7-10 August*, vol. I, pp. WeC3-3 – WeC3-9, International Society of Information Fusion (ISIF), 2001.
- [2] R. Kosara, S. Miksch, and H. Hauser, "Focus plus context taken literally," *IEEE Comp. Graph. and App.*, vol. 22, no. 1, pp. 22–29, 2002.
- [3] D. A. Keim, "Information visualization and visual data mining," *IEEE Trans. Vis. and Comp. Graph.*, vol. 8, no. 1, pp. 1–8, 2002.
- [4] M. G. Jones and S. G. Nikolov, "Region-enhanced volume visualization and navigation," in *Proc. SPIE*, vol. 3976, pp. 454–465, 2000.

- [5] M. G. Jones and S. G. Nikolov, "Volume visualisation via region-enhancement around an observer's fixation point," in *MEDSIP 2000 (International Conference on Advances in Medical Signal and Information Processing)*, Bristol, UK, 4-6 September, vol. 476, pp. 305–312, 2000.
- [6] United Bristol Healthcare NHS Trust, "Method and apparatus for displaying volumetric data," *British Patent Application No: 99124338.0, May 1999. US Patent Application No. 09/579,814. European Patent Application No. 00304381.7*, 1999.
- [7] The National Library of Medicine, "The Visible Human Project [Online]. http://www.nlm.nih.gov/research/visible/visible_human.html," 2003.
- [8] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," *Comp. Graph.*, vol. 21, pp. 163–169, 1987.
- [9] M. Levoy, "Display of surfaces from volume data," *IEEE Comp. Graph. and App.*, vol. 8, pp. 29–37, 1988.
- [10] A. Kaufman, D. Cohen, and R. Yagel, "Volume graphics," *Computer*, vol. 26, pp. 51–64, 1993.
- [11] M. Meißner, J. Huang, D. Bartz, K. Mueller, and R. Crawfis, "A practical evaluation of popular volume rendering algorithms," in *Proc. 2000 Symposium on Volume Visualization*, pp. 81–90, October 2000.
- [12] L. Westover, "Footprint evaluation for volume rendering," in *Proc. SIGGRAPH '90*, pp. 367–376, Aug. 1990.
- [13] M. Woo, J. Neider, T. Davis, D. Shreiner, and OpenGL Architecture Review Board, *OpenGL Programming Guide*. Reading, Massachusetts: Addison Wesley, 3rd ed., 1999.
- [14] H. J. Noordmans, H. T. M. van der Voort, and A. W. M. Smeulders, "Spectral volume rendering," *IEEE Trans. Vis. and Comp. Graph.*, vol. 6, no. 3, pp. 196–207, 2000.
- [15] E. B. Goldstein, *Sensation and Perception*. Pacific Grove, USA: Brooks/Cole, 4th ed., 1996.
- [16] R. Stokking, K. J. Zuiderveld, and M. A. Viergever, "Integrated volume visualization of functional image data and anatomical surfaces using normal fusion," *Human Brain Mapping*, vol. 12, no. 4, pp. 203–218, 2001.
- [17] P. Rheingans and D. Ebert, "Volume illustration: Non-photorealistic rendering of volume models," *IEEE Trans. Vis. and Comp. Graph.*, vol. 7, no. 3, pp. 253–264, 2001.
- [18] S. Saida and M. Ikeda, "Useful visual field size for pattern perception," *Perception and Psychophysics*, vol. 25, pp. 119–125, 1979.
- [19] M. L. Brady, K. K. Jung, H. T. Nguyen, and T. P. Nguyen, "Interactive Volume Navigation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 4, pp. 243–256, July–September 1998.
- [20] K. Myszkowski, P. Rokita, and T. Tawara, "Perception-based fast rendering and antialiasing of walkthrough sequences," *IEEE Trans. Vis. and Comp. Graph.*, vol. 6, no. 4, pp. 360–379, 2000.