To appear in *Proc. of the Fifth International Coference on Information Fusion, 2002.*

Fusion of Multi-Modality Volumetric Medical Imagery

Mario Aguilar
Knowledge Systems Laboratory
Mathematical, Computing and Information Sciences Department
Jacksonville State University
marioa@ksl.jsu.edu

Joshua R. New
Knowledge Systems Laboratory
Mathematical, Computing and Information Sciences Department
Jacksonville State University
newj@ksl.jsu.edu

Abstract - *Ongoing efforts at our laboratory have targeted the development of techniques for fusing medical imagery of various modalities (i.e. MRI, CT, PET, SPECT, etc.) into single image products. Past results have demonstrated the potential for user performance improvements and workload reduction. While these are positive results, a need exists to address the three-dimensional nature of most medical image data sets. In particular, image fusion of three-dimensional imagery (e.g. MRI slices) must account for information content not only within a given slice but also across adjacent slices. In this paper, we describe extensions made to our 2D image fusion system that utilize 3D convolution kernels to determine locally relevant fusion parameters. Representative examples are presented for fusion of MRI and SPECT imagery. We also present these examples in the context of a GUI platform under development aimed at improving user-computer interaction for exploration and mining of medical data.*

Keywords: Image fusion, sensor fusion, medical imaging, multi-dimensional visualization, decorrelation, 3D kernels.

1 Introduction

The widespread availability of medical imaging devices has led to an explosion in the amount of information that medical specialists must confront in diagnosing and planning treatment. To avoid situations where the imaging information can overwhelm medical personnel or can degrade patient care, methods for combining multi-source imaging information must be developed. At Fusion 2001, we presented our work [1] on utilizing a neuro-physiologically-motivated fusion architecture for combining multi-modality medical imagery. The approach is based on techniques originally developed for fusion of low-light visible and thermal infrared for a night-vision system [2].

While our original report presented positive results regarding fusion of 2-, 3-, and 4-modality medical imagery, the approach was not capable of accounting for volumetric information present in medical imaging modalities such as MRI and SPECT. In particular, these modalities are recorded as multi-plane (slice) images across the organ of interest. While it is natural to associate the individual images as being the main tool used by radiologists in diagnosis, it is also true that medical specialists rely heavily on the information available across adjacent slices. Hence, it is routine practice for radiologists to utilize multi-slice displays in order to gain volumetric contextual information and assess the extent of features in all three dimensions. Furthermore, we identified the need to perform image enhancement across slices so as to obtain better definition of tissue in the context of their extent in all three dimensions. The work reported here consists of our initial results in extending the neurophysiologcally-motivated fusion architecture to three dimensions.

1.1 Neurophysiologically-motivated fusion

The visual system of primates and humans contains three types of light sensors known as cones, which have overlapping sensitivities (short, medium, and long wavelengths). It is through the combination of these three sensor sets that we obtain our perception of color. Circuitry in the retina is functionally divided into two stages. The first one utilizes non-linear neural activations and lateral inhibition within bands to enhance and normalize the inputs. The second stage utilizes similar neural components in an arrangement of connections that lead to between-band competition that in turn produces a number of combinations of the three original bands [5]. This last stage of processing enhances the complementary information that exists in each of the bands (e.g. a spectral decorrelation operation).

The fusion architecture is motivated by this basic connectivity. Each of the processing stages is implemented via a non-linear neural network known as the *shunt* (described in the next section). The resulting non-linear combinations lead to information decorrelation not unlike what is usually targeted by principal component analysis techniques. However, in neuro-physiological systems and in our architecture, the non-linear operator has a very narrow spatial window providing a better-tuned decorrelation. In addition, the operator is modulated by more globally defined statistical characteristics of the input

that produce normalization, smoothing, and between-band calibration.

2 Methods

2.1 2D Shunt

The basic building block of the fusion architecture is a non-linear neural network known as a shunting neural network [4]. This neural network, which acts like a filter, models the dynamics of neural activation in response to three different signals: an excitatory one-to-one input corresponding to the current pixel, an inhibitory input from surrounding neurons, and passive activation decay (i.e. a leaky integrator). The expression that captures these interactions in a dynamical system is defined in terms of the following differential equation:

$$\dot{x}_{ij} = -Ax_{ij} + (B - x_{ij})[CI^C]_{ij} - (D + x_{ij})[G_S * I^S]_{ij} \quad (1)$$

Here, x_{ij} is the activation of each cell *ij* receiving input from each of the pixels in the input image. *A* is a decay rate, *B* and *D* are the maximum and minimum activation levels respectively and are set to 1 in the simulations. *C* and G_s, as represented by the following equation,

$$G(x, y) = \frac{1}{2\pi s^2} e^{-\frac{(x^2 + y^2)}{2s^2}} \quad (2)$$

serve to weigh the excitatory input (I^c) vs. lateral inhibitory (I^s) image inputs.

The neural network consists of a two dimensional array of these shunting neurons with dimensions *ij* corresponding to the width and height of the input image. When the input is applied, the network rapidly reaches equilibrium, which produces the resulting output image. This equilibrium state can be understood in terms of the possible values of *x* after the neuron has reached a steady state as shown in the following equation:

$$x_{ij} = \frac{[CI^C - G_S * I^S]^+_{ij}}{A + [CI^C + G_S * I^S]_{ij}} \quad (3)$$

Here, it is straightforward to understand the numerator as a contrast enhancement operation since it represents a difference of Gaussians. The denominator serves to normalize the activation of x_{ij} with respect to the activity of its neighborhood. In effect, the combination of these two operations leads to the dynamic range compression of the input image in conjunction with contrast enhancement. Parameter *A* serves to control the characteristics of the operator, from ratio processing (when *A* is small with respect to the local statistics) to linear filtering (when A is comparatively large).

In the case in which the operator is used to combine two bands, one of the bands is mapped to the center and the other to the surround. In the case where band 1 is mapped to the center, each of the pixels from band 1 is used to drive the excitatory input of its corresponding shunting operator. Then, a corresponding broader area of the image from band 2 is used as the surround input of the same shunt operator. The result is the contrast enhancement of information in band 1 as matched against band 2. The relationship between this operation and decorrelation has been previously documented [5].

2.2 3D Shunt

As discussed previously, modern medical imaging techniques such as MRI and SPECT produce multi-slice image volumes. For this reason, we have extended the shunting operator to three dimensions in order to account for information content between slices. Furthermore, it is desirable to process such image sets and perform image fusion while exploiting the information present in the third dimension. To accomplish this, it is necessary to modulate contrast enhancement and information decorrelation as a function of the local context in all three dimensions. The shunting equation that performs this function is as follows:

$$\dot{x}_{ijk} = -Ax_{ijk} + (B - x_{ijk})C[G_c * I^C]_{ijk} - (D + x_{ijk})[G_S * I^S]_{ijk} \quad (4)$$

Here, the activity of cell *x* (i.e. the resulting pixel value) is defined in terms of the contrast enhancement and dynamic-range normalization defined by a three-dimensional neighborhood around the cell. In this equation, we introduce two Gaussian terms: the first one modulating the excitatory inputs to cell x_{ijk}, and the second one modulating its inhibitory inputs. This extension required the use of a three-dimensional Gaussian kernel as follows:

$$G(x, y, z) = \frac{1}{4\pi^2 s^3} e^{-\frac{(x^2 + y^2 + z^2)}{2s^2}} \quad (5)$$

The combination of both kernels (G_c and G_s) in equation 4 leads to a three-dimensional operator that we represent by the following icon:

2.3 Color-opponency for image fusion

The basic fusion architecture consists of two distinct processing stages. In the first one, as in the retina, we utilize a shunting operation to obtain within-band image enhancement and normalization. This produces contrast enhancement, dynamic range calibration, and normalization of input images. In this version however, the input to the 3D shunting operator is a stack of image slices. Here, one slice is processed at a time with the shunt operating on the current slice as well as those slices immediately above and below. The result is a slice with enhanced contrast characteristics due to between-slice context information.

The second stage adopts the use of the same non-linear neural network operator to produce between-band decorrelation, information enhancement, and fusion. These stages for the two-band fusion case are illustrated in Figure 1. Here, the icon for the 3D shunting operator is colored coded so as to indicate the source of its inputs.

The shunt combinations of the second stage, as shown in Figure 1, provide two unique sets of information-rich images. These combinations perform paired decorrelations between the input image modalities. For instance, in the case of the top combination (red-center and green-surround in Fig. 1) the result is the decorrelation of the first modality (T1) from the second modality (T2). In other words, it enhances information that is present in band 1 but not in band 2. The resulting images from both decorrelations are then mapped to the red and blue channels of the color display. The enhanced T1 image was chosen as the input to the green channel in order to highlight information from that modality.

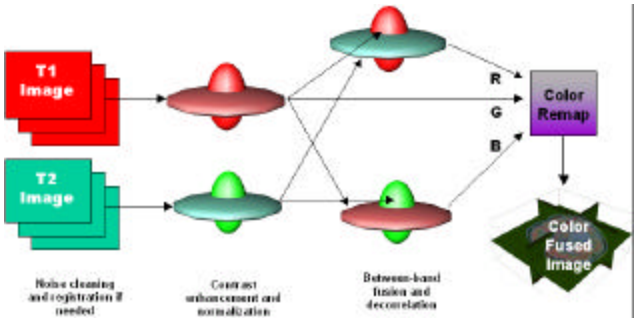

Figure 1. Two-band fusion architecture used for processing functional MRI imagery. Concentric spherical volumes represent a shunting neural network operator. Color coding indicates polarity of the operations. See Text for details.

Another processing stage may be introduced prior to producing the final color image that remaps the color assignments from those derived by the fusion process. Here, a mapping from RGB to HSV space allows the user to manipulate the appearance of the image (e.g. hue remap) to obtain a more natural coloring scheme. The modified HSV values are mapped back to RGB to be used in generating the final color fused image.

The architecture for three-band MRI fusion is illustrated in Figure 2. Here, the first stage of processing is as before, where each of the input bands is separately contrast enhanced and normalized. Then, two between-band shunting operations produce distinct fusion products. The first one decorrelates the information between bands 1 (MRI-T1) and 2 (MRI-PD). The second does it for bands 3 (MRI-T2) and 2. The resulting fused images are then mapped to the I and Q (also known as red-green and blue-yellow) components of the YIQ color space of the image. In this case, the resulting images highlight information that is unique to bands 1 and 3 in red and blue hues, while presenting unique information to band one in green and yellow hues. The Y or achromatic component is derived from the enhanced band 2 image that provides the most faithful structural details. The YIQ components are then mapped to RGB space.

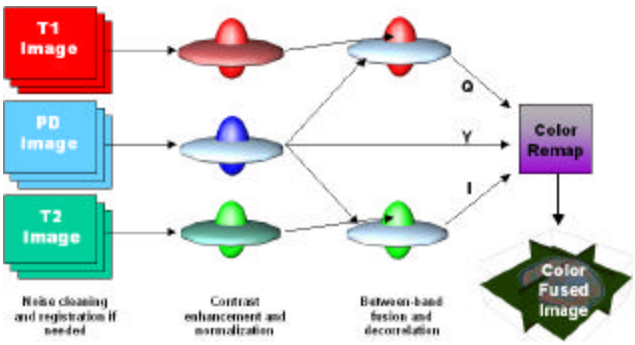

Figure 2. Three-band MRI fusion architecture. See text for details.

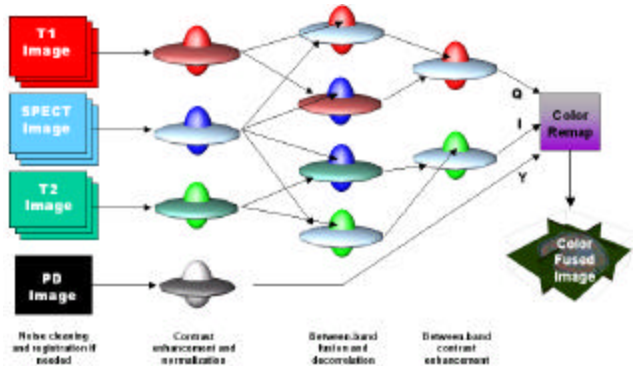

Figure 3. Four-band MRI/SPECT fusion architecture. See text for details.

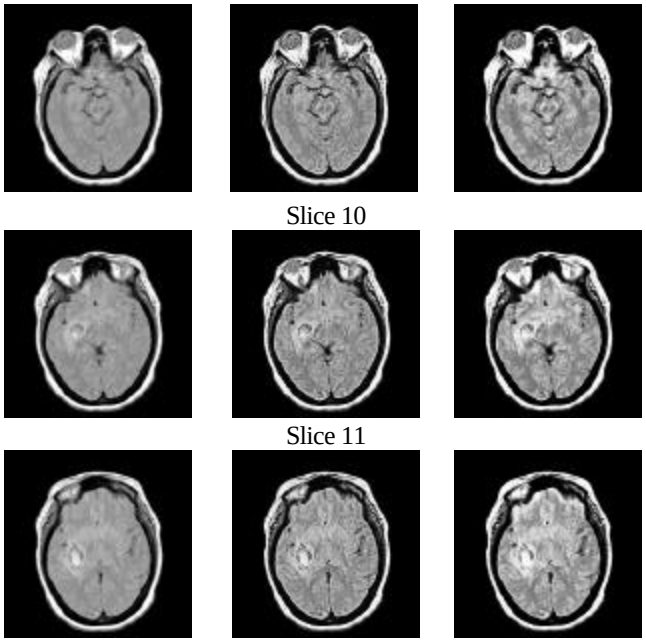

Figure 4. Comparison of 2D vs. 3D within-band shunting of three adjacent image slices. Left column: original unprocessed imagery. Middle column: 2D shunted imagery. Left column: 3D shunted imagery.

The architecture for four-band fusion is shown in Figure 3. Here, the second stage of processing produces the decorrelation between T1-weighted and SPECT, as well as between T2-weighted and SPECT. Notice that the decorrelation is done in both directions for each of the pairs. The most noticeable difference is the addition of a third processing stage. This additional between-band competition leads to further color contrast enhancement and decorrelation as suggested by connectivity in primary visual cortex in primates. The two resulting decorrelated images are mapped to the chromatic I and Q channels. Once again, to preserve the high resolution of the MRI imagery, the structural modality MRI-PD image is mapped to the Y or achromatic channel.

3 Results

The fusion architecture was applied to medical imagery obtained from the Brain Atlas web site (see [1] for reference). The imagery included 3-modality MRI and SPECT slices. We present the results of processing the imagery using the new shunting operator in three of the stages of the fusion process: within-band enhancement, between-band decorrelation, and color fusion.

3.1 Within-band shunting.

The 3D shunting operator (eq. 3) was applied to each of the 3 original MRI image modalities (PD, T1-weighted, and T2-weighted) (see [1] for a description of the modalities). The resulting contrast-enhancement and normalization were compared to the results obtained using the original 2D shunting operator. Figure 4 presents this comparison. We include the original unprocessed imagery to provide a sense as to the success of both methods in bringing out significant amount of information. Here, the parameters of both operators were the same for the two-dimensional extent of the neighborhoods. In addition, the three-dimensional extent of the neighborhood for the 3D shunt was set to two slices above and two below. We have experimented with different shape Gaussians for the third dimension and have found that a symmetrical neighborhood works well for this data set. We anticipate that as resolution along the third dimension degrades (i.e. distance between slices) the Gaussian kernel would need to be made narrower accordingly. We hope to test this idea as soon as we obtain a relevant image set.

3.2 Between-band shunting

The 3D shunting operator was again applied across slice stacks, this time utilizing one modality as the input to the center and the other to the surround of the operator. The result is a decorrelation of information between the bands as explained earlier. Furthermore, this decorrelation includes a modulation factor derived from the context of adjacent slices. The results for 2-band fusion are shown in Figure 5.

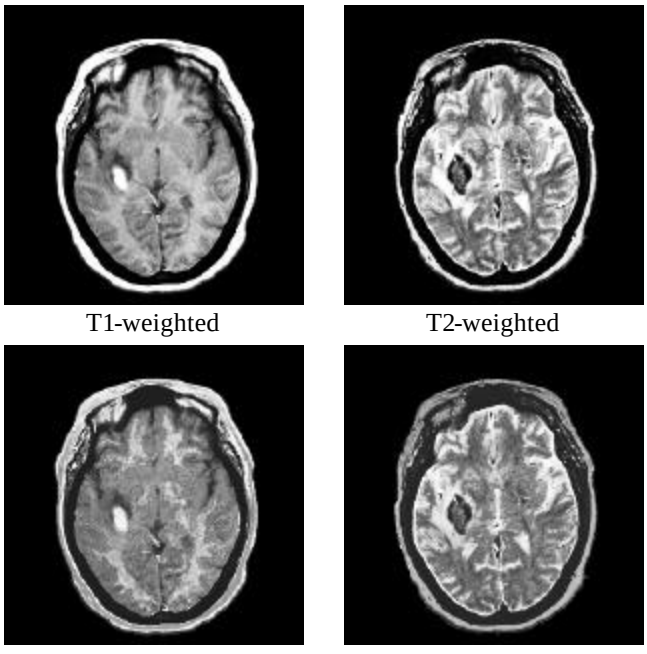

Figure 5. Results of applying a 3D shunt to combine and decorrelate T1-weighted and T2-weighted images.

3.3 Color-fusion of 2-, 3-, and 4-modality imagery

The imagery was further processed using the 3D fusion architecture to produce the color-fused results that combine information from all input modalities. Here, both color and brightness characteristics of the resulting image are used to convey information about the characteristics and source of the information in the original imagery. Figure 6 presents side-by-side the result of fusing 2- and 3-band MRI imagery. Three slices are shown to provide across-slice context and highlight the efficacy of the method in enhancing information across all three dimensions.

In figure 7, we present a comparison of the original and updated 4-band fusion results as applied to slice 11. Here, the result on the left is as obtained using the original fusion architecture. On the right, the result of the 3D fusion presents enhanced information content highlighted by the enhancement produced across slices.

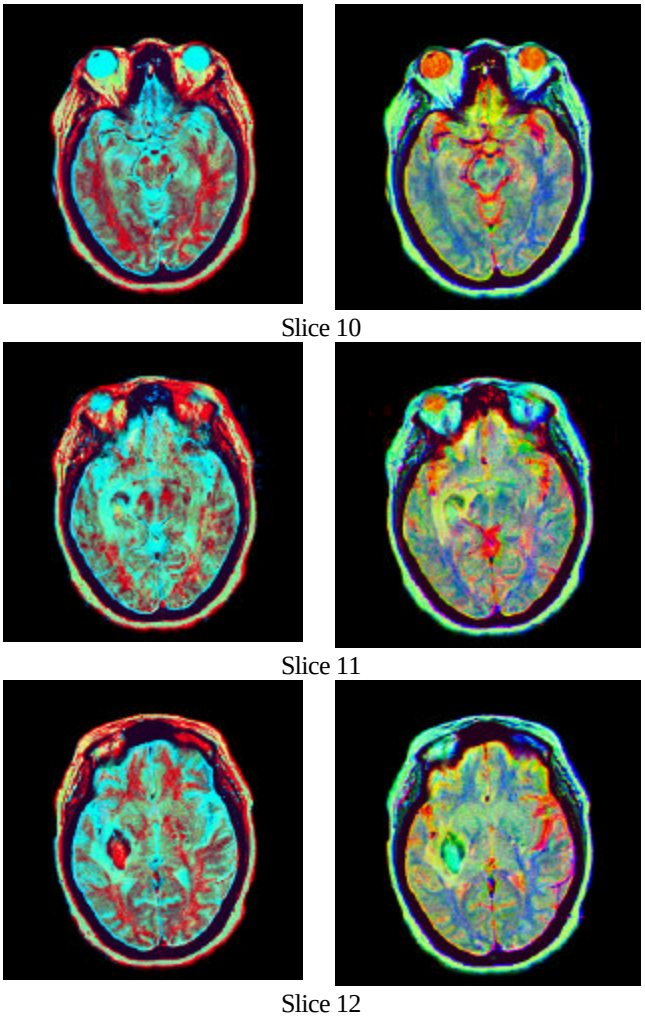

Figure 6. Comparison of 2- and 3-band fusion results obtained on the MRI imagery. Left column : 2-band fusion of T1- and T2-weighted slices. Right column : 3-band fusion of PD, T1- and T2-weighted slices.

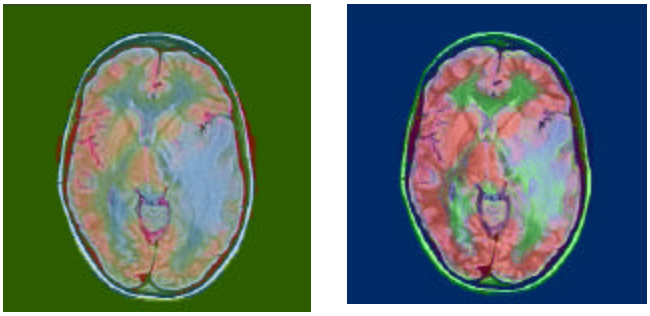

Figure 7. Comparison of 2D vs. 3D 4-band fusion of MRI and SPECT. Notice the enhanced delineation of SPECT information as indicated by its red hues.

Finally, to emphasize the benefits of the 3D operator in enhancing information across slices, Figure 8 presents a comparison of 3D imagery as obtained with the new fusion architecture (on the left) and the original architecture (on the righ). Here, we have utilized our 3D visualization tools to recreate three orientation planes and their corresponding fused imagery overlaid on them. These orientations (coronal, saggital, and axial) represent orthogonal views through the cranial volume. The results of the 3D fusion architecture demonstrate better definition of the highlighted volume (tumor) in all three dimensions.

4 Conclusions

We have extended our fusion architecture to address the three-dimensional nature of medical image sets. In particular, we have introduced a 3D shunting operator that contrast enhances and normalizes information present in the immediate neighborhood of each pixel including adjacent image slices. Our results compare favorably against the results of the original 2D fusion architecture. Better definition of image details and volumetric information make this approach a significant advance in medical image fusion. Ongoing efforts will help us assess the significance of the information gains and fine tune our fusion architecture.

Acknowledgements

This work was supported by a Faculty Research Grant awarded to the first author by the faculty research committee and Jacksonville State University. Opinions, interpretations and conclusions are those of the authors and not necessarily endorsed by the committee or Jacksonville State University.

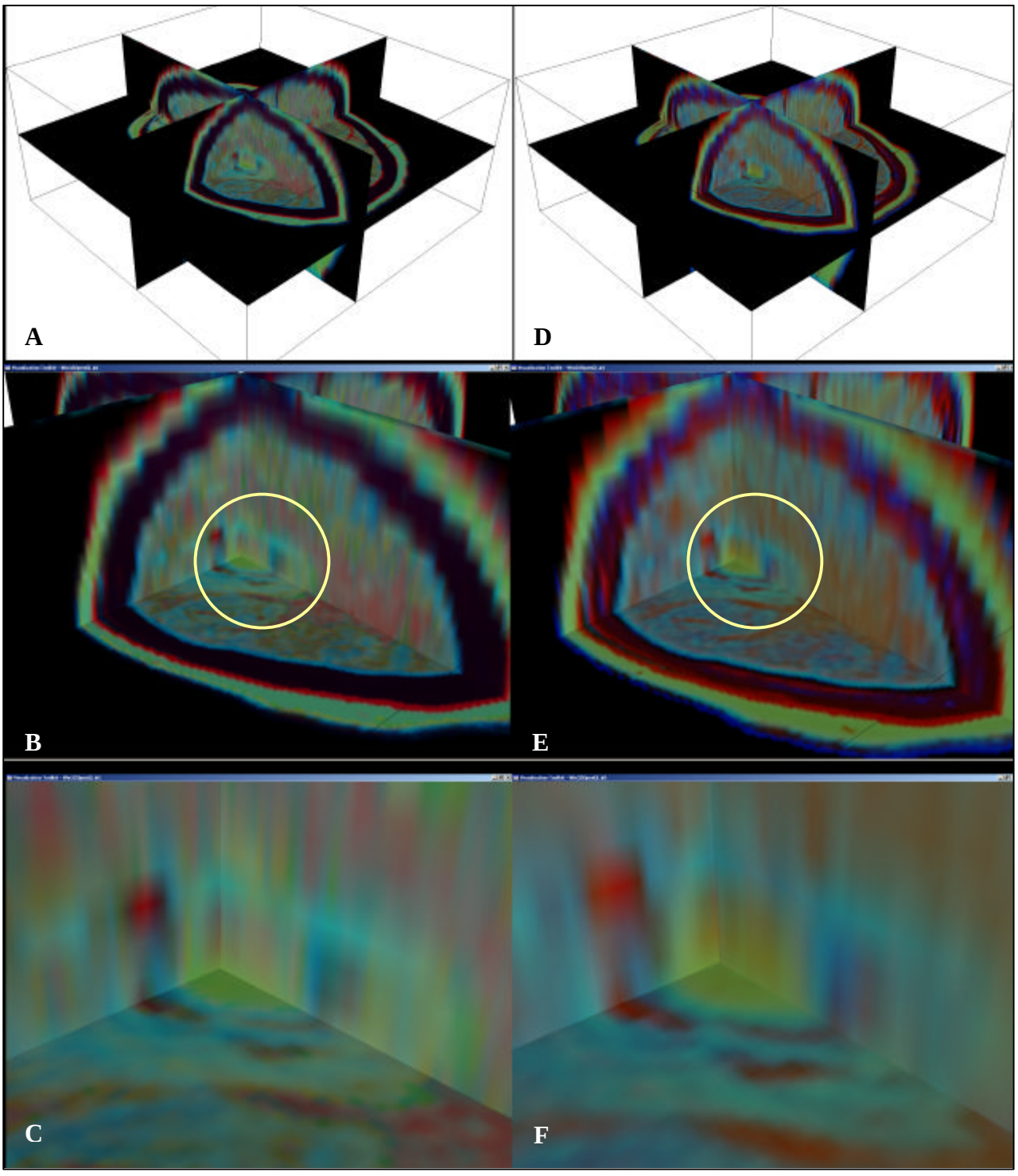

Figure 8. Three-dimensional visualization of 3-band fusion results. Left column presents results obtained using the new 3D shunting operator. Images on the right column are the results obtained using the original shunt operator. A. Far view of the three orthogonal image planes obtained with the 3D fusion architecture. B. Closer view of the intersection of the planes highlighting the volume of interest (tumor). C. Closest view of the area of interest where the contrast-enhancement across slices can be observed (slices change in the up-down direction). D. Far view of the three orthogonal image planes obtained with the 2D fusion architecture. E. Closer view of the intersection of the planes highlighting the volume of interest (tumor). F. Closest view of the area of interest where the contrast-enhancement across slices is noticeably absent.

References

[1] M. Aguilar and A.L. Garrett, "Neurophysiologically-motivated sensor fusion for visualization and characterization of medical imagery", *Proc. of the Fourth International Conference on Information Fusion*, 2001.

[2] M. Aguilar, D.A. Fay, W.D. Ross, A.M. Waxman, D.B. Ireland, and J.P. Racamato, "Real-time fusion of low-light CCD and uncooled IR imagery for color night vision", *Proc. Of SPIE Conf. On Enhanced and Synthetic Vision*, **3364**, 1998.

[3] P. Schiller and N.K. Logothetis, "The color-opponent and broad-band channels of the primate visual system", *Trends in Neuroscience*, **13**, pp.392-398, 1990.

[4] S. Grossberg, Neural Networks and Natural Intelligence, Cambridge, MA: MIT Press, 1988.

[5] M. Aguilar and A.M. Waxman, "Comparison of opponent-color neural processing and principal components analysis in the fusion of visible and thermal IR imagery." *Proc. of the Vision, Recognition, and Action: Neural Models of Mind and Machine Conference*, Boston, MA, 1997.q