

MED-LIFE: A DIAGNOSTIC AID FOR MEDICAL IMAGERY

Joshua R. New, Erion Hasanbelliu and Mario Aguilar
Knowledge Systems Laboratory, MCIS Department
Jacksonville State University, Jacksonville, AL

ABSTRACT

We present a system known as Med-LIFE (Medical application of Learning, Image Fusion, and Exploration) currently under development for medical image analysis. This pipelined system contains three processing stages that make possible multi-modality image fusion, learning-based segmentation, and exploration of these results. The fusion stage supports the combination of multi-modal medical images into a single, color image while preserving information present in the original, single-modality images. The learning stage allows experts to define the pattern recognition task by interactively training the system to recognize objects of interest. The exploration stage embeds the results of the previous stages within a 3D model of the patient's skull in order to provide spatial context while utilizing gesture recognition as a natural means of interaction.

1. INTRODUCTION

As more powerful imaging techniques grow ever more pervasive, medical experts often find themselves overwhelmed by the large number of images being produced. In addition, many of these images are significantly complementary and often lead to an increase in workload. Experts typically have to view and analyze multiple images which force them to follow a tedious scanning procedure. Inadvertently, this work overload leads to a decrease both in the quantity and quality of healthcare that can be provided.

The first component, an image fusion architecture, was implemented to combine and enhance information content from multiple medical modalities. This fusion architecture takes advantage of the volumetric nature of the imagery in order to better contrast enhance and de-correlate the information present in single images before combining them into a single color image. By providing a single fused image, this architecture can improve the speed at which experts can make a diagnosis decision.

The second component, a learning system, has been developed to provide an interface for training the computer to perform automated segmentation and preliminary diagnosis. Instead of focusing our efforts on developing a general learning engine, we are pursuing a targeted learning design whereby users help define the context and constraints of a specific recognition task. This is accomplished by allowing the user to define features or areas of interest in the imagery. This information is then utilized by a machine-learning system to establish an appropriate mapping between inputs and desired landmark identity. In other words, the system establishes input-based recognition patterns that uniquely identify the characteristics of interest to the user. Furthermore, this task-specific recognition pattern can be encapsulated as an autonomous agent that can in turn be used to identify and flag/highlight areas of interest that may be present in other areas of the image or within large patient databases. This system, we hope, could further improve the amount of care that can be provided by aiding in the process of pre-diagnosis or thoroughness of care by checking other available database images for non-diagnosed areas of concern.

The third component, an interactive system for the visualization of fusion and pattern recognition results, was developed for exploring the results in both two and three dimensions. This system utilizes the inherent three-dimensional information of the original imagery to create volumes for a more intuitive presentation and interaction with the user. This allows for localization of the task-relevant image features and planning of surgeries. In addition, a real-time gesture recognition interface was developed for users to intuitively navigate through the data. The interface developed allows users to interact with a 3D model of the skull simply by moving or posing their hand in front of a camera. The exploration component provides the key to successful visualization and information analysis by allowing users to quickly and easily understand the information generated by the other components of the system.

2. THE MED-LIFE APPLICATION

A central theme of the Med-LIFE system is as a true human-computer system in which the expertise of the user/radiologist can be leveraged to enhance the performance of the overall system. This is made possible by allowing the user to continuously monitor and modify the processes involved in fusing and segmenting the information. In this way, Med-

LIFE is more than just a software product or diagnostic aid, but a system that can capture and exploit the combined capabilities of its users as well as the proficiencies of the computer system.

The Med-LIFE GUI was developed using QT in order to be platform independent. The functionality of Med-LIFE was implemented primarily in C++ with the VTK, IPL, and OpenCV libraries. Med-LIFE is a pipeline system consisting of three processing stages associated with the components described in the previous section: fusion, learning, and exploration. Each stage is implemented as a tab in a graphical-user interface which assists the user in understanding and organizing the workflow. This follows a logical order of allowing fusion of image modalities, followed by computer segmentation of the identified fused results, and then exploration of the learned results in that order.

In the remaining sections, we describe each of the components' theoretical foundations and corresponding implementation. The discussion in each section is followed by a description of the software module as implemented in the final system. For the purposes of demonstrating the performance of the system, we illustrate the description with real cases extracted from a publicly available image database. Images used in Med-LIFE were spatially registered by the authors of Harvard's Whole Brain Atlas [1]. The modalities used typically consist of three morphological modalities (PD, T1, and T2) and one functional (SPECT) modality. These modalities were chosen in order to maximize both the morphological and functional information with the fusion architecture; however, the fusion architecture could be applied to any set of image modalities.

3. IMAGE FUSION STAGE

A neurophysiologically-based fusion architecture has been established for the combination of multi-modal medical imagery. The architecture is based on the visual system of primates which is itself involved in performing image fusion in order to obtain color perception. Information-preserving fusion is obtained through two processes. First, non-linear neural activations and lateral inhibition within bands enhance and normalize the inputs. Second, similar neural components perform between-band competition to spectrally de-correlate the information which produces a number of combinations of the three original bands.

The processing stages described above are implemented via a non-linear neural network known as the shunt operator [2]. This shunt operator has been extended to a 3D kernel in order to take into account the volumetric nature of modern MR imagery [3]. Multiple fusion architectures have been tried and the hybrid architecture shown in Fig. 1 has been found to yield the most effective 4-band fusion results [4].

In Fig. 1, the T1, SPECT, T2, and PD modalities of a given case are filtered through a series of 2D and 3D shunt operators into Y, I, and Q chromatic channels which were then combined to form a single, color image. While only the combinations shown in Fig. 1 are used for creating the color-fused image, all valid single-opponent combinations are created. The data from this plethora of images is used by the next stage of the system pipeline.

A screenshot of the Fusion tab is shown in Fig. 2. Group A displays the result of the fusion process for the slice-of-interest, which can additionally be zoomed and panned with the mouse to permit more thorough inspection. Group B is the slice slider which allows selection of a slice-of-interest. Group C allows the user to select the type of fusion result to visualize. Group D is a scrollable text box which displays case-relevant information. Group E contains the original images used in the fusion process. Group F provides the ability to swap the fusion result with the respective image so that the MR imagery can be more thoroughly inspected.

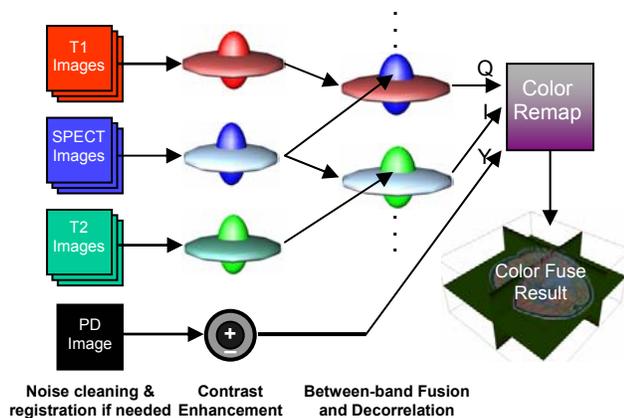


Figure 1. Default 4-Band Fusion Architecture

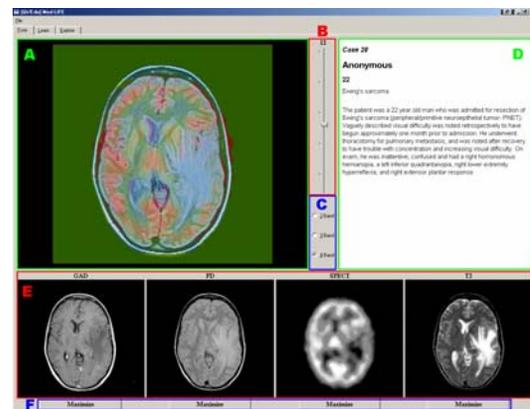


Figure 2. Image Fusion Tab (demonstrating a 4-band fusion example)

4. LEARNING STAGE

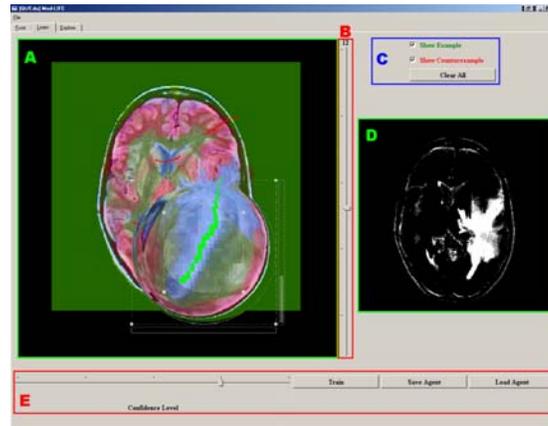
The multitude of images created at the image fusion stage serve as input features for a learning system to be used for image segmentation. A neural network architecture known as ARTMAP has been used to provide a supervised, incremental, nonlinear, fast, stable, online/interactive learning system [5] to assist interactive diagnosis. Learning is accomplished by allowing the user to interactively train the computer to recognize areas of interest (such as tumor regions). Feedback via a mechanism for highlighting similar areas found in the current slice, adjacent slices, or even other patients helps the user monitor the learning process. Straightforward codification of the user's task of interest into robust AI agents is allowed by leveraging off the expert's knowledge. These agents can later be loaded to pre-screen images by highlighting areas of potential interest or scouring through a database of patient images.

In order to provide robust segmentation across slices and patients, several methods were used. First, a confidence measure is used so as to generalize the quality of segmentation across slices. Second, a heterogeneous network of SFAM voters was established that varies the learning system parameters [6]. Third, the user is brought into the loop for training and correcting the system. By doing so, the user can interactively adapt an agent toward better overall performance.

The selection of areas of interest is a task requiring high precision and requires the user to zoom in to differentiate between targets and non-targets at the pixel level (see green and red markings in Fig. 3, group A). However, it is important for the user to maintain spatial context; that is, to know where in the image one is currently looking. In order to allow magnification while preserving spatial context, contextual zooming was implemented with IDELIX's PDT SDK [7].

The effectiveness of the learning stage is maximized due to the richness of the input features, the strength of the learning system, and the inclusion of the user within the learning loop. An example of the learning stage's contextual zoom and segmentation results can be seen in Fig. 3.

Group A consists of the fuse result selected in the previous pipeline stage. Group B is the slice slider that allows traversal of all available slices. Group C consists of checkboxes that allow for the customized viewing of examples and counterexamples in group A. Group D shows the segmentation results from the SFAM voters after training on five swipes of the mouse in Group A. Group E consists of agent interface tools that allow the customization, training, saving, and loading of AI agents. When an agent is trained or loaded, a transparent overlay of group D can be used in group A to denote regions of interest.



**Figure 3. Learning Tab
(demonstrating training and segmentation of carcinoma tissue)**

5. EXPLORATION STAGE

The purpose of the exploration stage is to provide a natural means of interaction between the user and all the data used and generated in the previous stages. To do this, the original MR images are provided since users may want to refer back to the original imagery in certain cases. The 2D fusion images for the current slice, the slice above and the slice below are provided along the bottom to aid in contextual slice navigation; that is, to allow the users to easily determine and follow the extent of any area of interest throughout the cranial volume.

In order to facilitate natural navigation and understanding of the data, some additional capabilities were developed. First, the 2D fusion slices from the patient are imbedded within a 3D, patient-specific skull which is computed from the patient's segmented MRI imagery. The user can rotate, pan, and zoom in/out on this 3D object using the mouse. Second, blood flow information (such as that found in SPECT) is important for diagnosis, so a mechanism was developed to allow the user to customize the amount of SPECT overlaid on the current 2D fusion image.

In order to reduce the complexity of interaction between humans and computers, we have investigated the use of gesture recognition to support natural user interaction while providing rich information content. The system was inspired by Kjeldsen’s thesis which provides strong motivation for natural gesticulation as an interface modality. Technical details of the gesture recognition system can be found in [8,9]. The gesture recognition system developed utilizes only common hardware and software components to provide real-time recognition of hand location and number of fingers from a 640x480 camera feed. This information is then used to manipulate the cranial volume as shown in Fig. 4.



Figure 4. Gesture Interface Subsystem

A screenshot of the Exploration tab is shown in Fig. 5. Group A contains the generated skull with the slice of interest embedded within for contextual navigation. The embedded slice consists of the desired fusion image combined with a SPECT overlay and a transparent overlay of segmentation/recognition results. This volume can intuitively be rotated, panned, or zoomed using gesture recognition. Group B contains the usual slice slider used to update groups A, E, and F. Group C consists of the SPECT transparency slider which allows customization of the SPECT overlay in group A so that the desired amount of metabolic information may be displayed. Group D consists of the learning stage’s viewing options. Here the user may select the desired fusion result for display in groups A and F, remove the skull if it is not needed or hinders the view, remove the raw images in group E, or remove the SPECT overlay in group A. Group E consists of the original images for the slice of interest. Group F consists of the fusion slice of interest as well as the slice above and below for contextual slice navigation.

6. CONCLUSIONS

We have presented a human-computer system which highlights effective user interaction and information understanding. This is accomplished by exploiting useful information pre-processing techniques based on neuro-computational principles. In addition, the user is introduced as part of the processing loop to alter, improve, and benefit from the automated processing of the imagery. In future work, we hope to initiate the validation of the overall system and assess its usability. Furthermore, we are expanding the current capabilities of the system in two important areas: extended gesture recognition and enhanced volume manipulation.

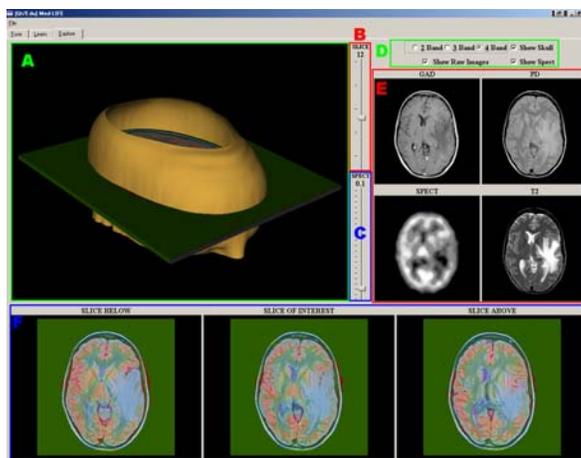


Figure 5. Exploration Tab

7. REFERENCES

- [1] K. Johnson and J. Becker, Whole Brain Atlas, <http://www.med.harvard.edu/AANLIB/home.html>, 1999.
- [2] S. Grossberg, Neural Networks and Natural Intelligence, Cambridge, MA: MIT Press, 1998.
- [3] M. Aguilar and J.R. New, "Fusion of Multi-Modality Volumetric Medical Imagery," *Proceedings of the 5th International Conference on Information Fusion*, Baltimore, MA, 2002.
- [4] M. Aguilar, J.R. New, and E. Hasanbelliu, "Advances in the Use of Neurophysiologically-based Fusion for Visualization and Pattern Recognition of Medical Imagery" *Proceedings of the 6th International Conference on Information Fusion*, Australia, 2003.
- [5] G.A. Carpenter, S. Grossberg, J.H. Markuson, J.H. Reynolds, and D.B. Rosen, "Fuzzy ARTMAP: A neural architecture for incremental supervised learning of analog multidimensional maps," *IEEE Transactions on Neural Networks*, 1992.
- [6] W. Streilein, A. Waxman, W. Ross, F. Liu, M. Braun, D. Fay, P. Harmon, and C.H. Read, "Fused multi-sensor image mining for feature foundation data," *Proceedings of 3rd International Conf. on Information Fusion*, Paris, France, 2000.
- [7] IDELIX, "MedLife—A PDT Integration Example : PDT in Medical Motion", Vancouver, BC, IDELIX Software Inc., <http://www.idelix.com/newsletters/newsletter20030807web.html#medlife>, 2003.
- [8] J.R. New, "A Method for Hand Gesture Recognition," *ACM Mid-SE Fall Conference*, Gatlinburg, TN, 2002.
- [9] J.R. New, E. Hasanbelliu, and M. Aguilar "Facilitating User Interaction with Complex Systems via Hand Gesture Recognition," *2003 SE ACM Conference*, Savannah, GA, 2003.