Registration and Fusion with Mutual Information for Information-preserved Multimodal Visualization

Augusto Cavalcante Valente and Wu, Shin-Ting
School of Electrical and Computer Engineering
University of Campinas, Campinas, Brazil
{avalente,ting}@dca.fee.unicamp.br

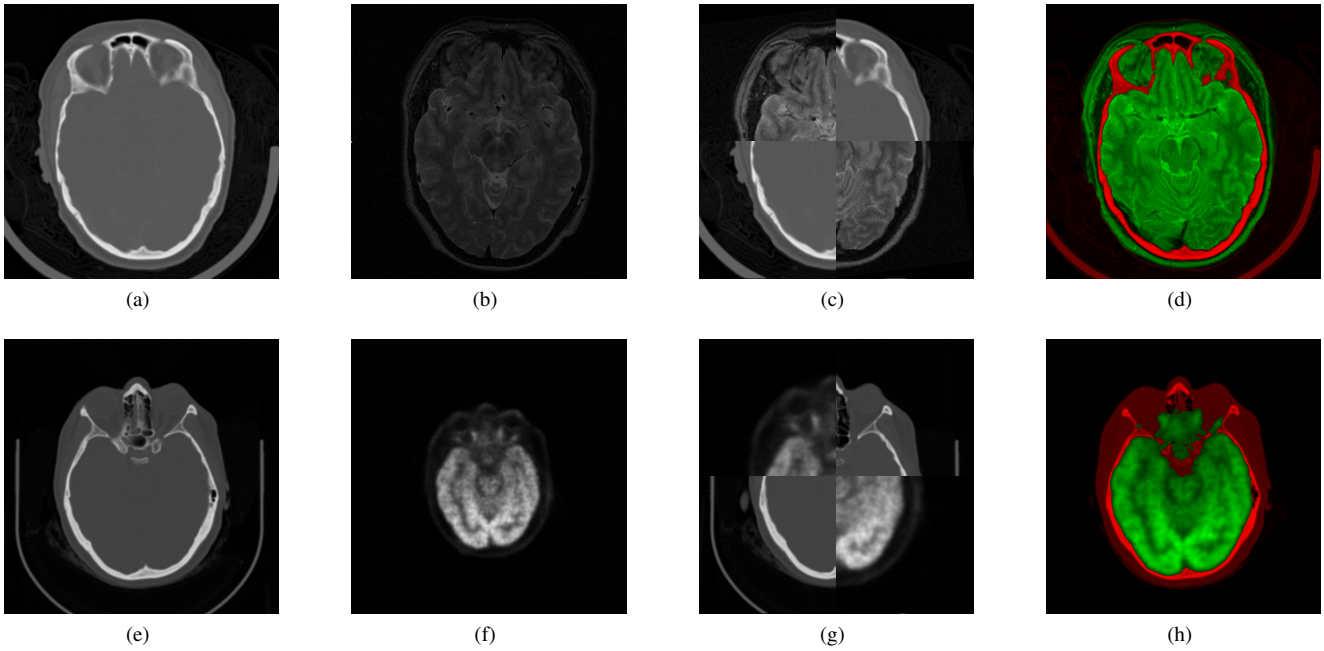


Fig. 1. a) CT and b) MR images of a patient, and their automatic c) registration and d) fusion (CT and MR voxels on red and green channels respectively). Also, e) CT and f) PET images of another patient and their automatic g) registration and h) fusion (CT in red and PET in green).

***Abstract*—Mutual information from Information Theory has been found to be quite effective in medical image registration. The efforts of neuro-scientists in specializing this measure in specific information measures have been recognized by Bramon et al. and they show that the specific information measures constitute a promising tool for choosing more informative data for each voxel in the fusion of images from different modalities. Aiming at information-preserved multimodal visualization, we present a proposal[1] to align and to combine multimodal images with a unified mutual information framework. As outcome, a control volume, which identifies the origin of each voxel, is generated and can be used e.g. for GPU-based volume rendering. Our experiments with CT, MRI and PET images modalities confirm the potential of our proposal to pre-process data for multimodal visualization.**

***Keywords*-Specific Information; Mutual Information; 3D Medical Image Fusion; 3D Medical Image Registration; Multimodal Visualization.**

[1]Master's Thesis

I. Introduction

In many image based medical studies, it is interesting to evaluate data acquired by different technologies or with varying parameters. This is because they provide complementary information useful for diagnosis or surgical planning. For example, X-ray based computed tomography (CT) is able to reproduce bone structures in detail, while magnetic resonance (MR) provides good contrast and discrimination of soft tissue. Positron emission tomography (PET) from nuclear medicine indicates areas where metabolism is higher, consisting of a functional image. Many other modalities are available, and clinicians are trained to examine all this information as separate images, mentally reconstructing spatial correspondences, a challenging task that can also often incur in precision errors.

Over the past decades, advances in acquisition technologies and digital image processing have been providing clinicians with better material and assistive tools to perform routine tasks. One of the possibilities is to simultaneously visualize different datasets in 3D, in a way that the different modalities

in use can still be distinguished. This is the goal of multimodal volume visualization, a volume rendering technique that has been actively researched [1]–[4]. To successfully carry out this visualization of distinct modalities, the 3D images need to be aligned in the same spatial domain, and also a fusion criterion has to be defined. This work focuses on both these steps.

It is necessary to spatially align the images because of different patient positioning in each acquisition. We focus on images of the head, therefore rigid alignments are suficient due to the neurocranium structure. This can be automated with an image registration algorithm. Among the variety of techniques that have been proposed, maximization of mutual information (MMI) has received much attention for effectively handling multimodal images without preprocessing or segmentation steps [5]. Recently, researchers have applied the same information theoretic approach to define multimodal data fusion criteria [6], [7]. This suggests the possibility of better integrating both registration and fusion steps.

Our work is motivated by the need to both register and fuse a pair of 3D images before sending them to the graphics pipeline for rendering. Differently from previous works, we present a unified architecture for registration and fusion on top of the same mutual information background, so that the outcome is a 3D information-preserved fused volume data.

II. Related work

The problem of integrating information from multiple image modalities acquired from the same patient is handled by performing two independent methods [8]: *registration* and *fusion*.

Diverse more or less automated approaches for *image registration* have been proposed for medical image registration [9]. For example, some are based on fiducial marker correspondence, which cannot be applied retrospectively. Others use geometrical features extracted from the images either interactively or automatically, and thus rely on the accuracy of this preprocessing task. A last class of methods aims at optimizing similarity measures computed directly from image intensities, which can be computationally expensive but more robust. Since the seminal works conducted in parallel by Collignon *et al.* [10] and Viola and Wells [11], the use of mutual information (MI) as a similarity measure for intensity based registration has become increasingly popular [5]. It is well suited for multimodal inputs because it assumes only a statistical relationship between intensities from different images, whereas e.g. cross correlation assumes a linear relationship. A series of works has been dedicated to mitigate its limitations w.r.t optimization [12] and extend it to consider spatial correlation [13]. We calculate mutual information using joint histograms as proposed by Collignon in order to reuse it on the fusion stage, as it is detailed in Section IV-B.

There is also a wide range of proposals for image fusion techniques. Wang *et al.* provide [14] an extensive review and a comprehensive unifying framework in the context of satellite imaging. Generally, some methods perform the combination in the spatial domain and others in a transform domain (e.g. using wavelets). Recently, Bramon *et al.* [7] propose to simply compose the fused image with selected voxels from the input sets, preserving original details and presenting unaltered information to clinicians. The voxel by voxel classification relies on a measure that is a decomposition of mutual information called *specific information*.

Most works deal with the registration and fusion stages separately, but the availability of techniques that have a common information theoretic background leads us to propose a unified procedure for generating a fused volume data from multiple single-valued volume data. The benefits are the reuse of calculation steps in the process and less intermediary data resampling.

III. Technical background

In this section, we summarize some important concepts that are necessary to understand our proposed paradigm.

A. Mutual information

Originated from the Information Theory field, mutual information (MI) is a measure of how much information a random variable (RV) contains about another [15]. In the context of image registration and fusion, the RVs are the image intensities. By definition, the mutual information I between two RVs A and B is

$$I(A,B) = \sum_{a,b} p_{AB}(a,b) \log \frac{p_{AB}(a,b)}{p_A(a)p_B(b)} \tag{1}$$

where $p_A(a)$ and $p_B(b)$ are the marginal probability distributions of A and B, and $p_{AB}(a,b)$ the joint distribution. Should A and B be independent, $p_A(a)p_B(b) = p_{AB}(a,b)$ and $I(A,B) = 0$. Therefore, MI measures the mutual dependency between the RVs.

Mutual information can also be expressed in terms of entropy, another basic concept from Information Theory. The entropy H of a RV measures the uncertainty associated with it, which can be interpreted as a measure of information. It also represents the dispersion of the probability distribution. We use the Shannon entropy definition of $H(A) = -\sum_a p_A(a) \log p_A(a)$, for a given RV A. The MI between two RVs A and B can be expressed as either

$$I(A,B) = H(A) + H(B) - H(A,B) \tag{2}$$

or

$$I(A,B) = H(B) - H(B|A) \tag{3}$$

where $H(A,B)$ and $H(B|A)$ are the joint and conditional entropies, based on the joint and conditional distributions, respectively. Eq. (3) expresses the decrease in uncertainty in B once A is known, and Eq. (2) illustrates that MI increases as the joint distribution dispersion decreases.

Among other properties, mutual information is symmetric ($I(A,B) = I(B,A)$) and nonnegative ($I(A,B) > 0$).

B. Image registration

The determination of a geometric transformation that maps corresponding points between the images may be formulated as an optimization problem

$$\hat{\theta} = \arg\max_{\theta} J(\boldsymbol{M}(\mathbf{x}), \boldsymbol{F}(T_\theta \mathbf{x})) \tag{4}$$

where $\boldsymbol{M}$ is a moving image (whose points are transformed), $\boldsymbol{F}$ is a fixed image, T is a geometric transformation with parameters θ applied to a point $\mathbf{x}$, and J is a cost function that measures the alignment. The estimated $\hat{\theta}$ parameters define the geometric transformation that we look for. Our work is restrained to rigid transformations, which have 6 parameters: 3 translations and 3 rotations.

Mutual information is commonly used as a similarity cost function when multimodal images are used, giving rise to the maximization of mutual information criterion (MMI) [10]. This criterion postulates that there will be maximum dependency between images when their intensity values are correctly aligned. Its validity comes from the fact that if the same material is observed by two different acquisition techniques, there will be some dependency between the measurements. Given the same sample, the knowledge of its value in an image provides insight about its value in another image.

C. Specific Information Measures

Mutual information only measures the average information one obtains about one RV A from observing another RV B, or vice-versa. Nevertheless, in some neuro-scientific applications, it would be interesting to know the information gained from a specific observation, or *specific information* to a particular event a_i, such that the average information gained over all possible events b_i is equal to mutual information. In this way, we may tell which events are more semantically important than others. Indeed, a specific information is derived from the decomposition of the mutual information, measuring the contribution of each outcome of an RV to the information gained about another RV.

There exists a variety of attempts to measure the specific information. In [7] three information-based measures have been applied to make informativeness comparison between the intensity values a and b assigned to the same voxel:

- Surprise or Specific Surprise (I_1): it can be viewed as a measure of the contribution of each intensity value b to the total information. Observations of b_i changes the probability distribution of the image A from $p_A(a)$ to $p_{A|B}(a|b)$ [16]

$$I_1(b, A) = \sum_a p_{A|B}(a|b) \log \frac{p_{A|B}(a|b)}{p_A(a)}. \tag{5}$$

- Predictability or Specific Information (I_2): it amounts to the reduction in uncertainty about A when one observes b [16]

$$\begin{aligned} I_2(b, A) &= H(A) - H(A|b) \\ &= -\sum_a p_A(a) \log p_A(a) \\ &\quad + \sum_a p_{A|B}(a|b) \log p_{A|B}(a|b). \end{aligned} \tag{6}$$

- Entanglement or SSI (I_3): it is the average reduction in uncertainty about A when one observes b [17]

$$I_3(b, A) = \sum_a p_{A|B}(a|b) I_2(a, B). \tag{7}$$

These measures are used in the fusion stage to choose the voxels whose intensities carry more information than others.

IV. Technique

Fig. 2 gives an overview of the control flow of our proposed technique. We start out with a pair of image volumes as input and an initial guess of their alignment, denoted by transformation T_0. The mutual information between these volumes is estimated from their joint histogram, and maximized w.r.t. a geometric transformation T_i by an iterative search algorithm. When the search converges, the volumes are said to be registered.

After registration, we select between corresponding voxels which one is more relevant for a fused dataset by a criterion based on specific information. This classification is stored on a control volume, which can be used to render a combination of the input volumes preserving their original intensities.

The techniques chosen for the registration and fusion stages share the same information theoretic background. This enables us to reuse calculations along the pipeline and to better integrate the process by avoiding unnecessary resampling steps. The main piece of calculation that gets reused is the joint histogram.

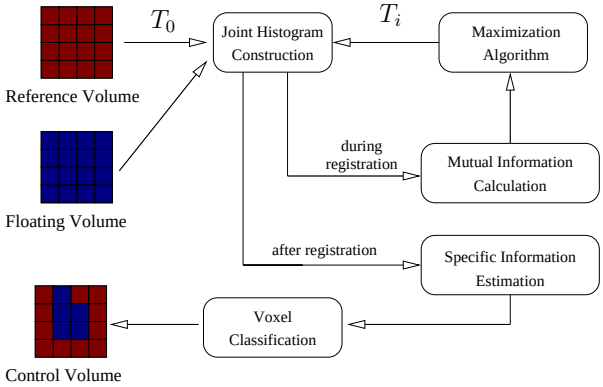


Fig. 2. Overview of the techniques applied to process input volumes into a fused dataset.

The joint histogram is the basis for the calculation of the mutual information metric used in registration and the specific information measures used in the fusion criterion. We now

discuss its construction algorithm and its application in the optimization process. We then detail the measures and criteria used in fusion that are able to reuse it.

A. Joint histogram and MI calculation

The fundamental part of calculating MI is the estimation of the probability densities, in particular p_{AB}. We adopt the proposal by Collignon [10] to estimate it using joint histograms. They are constructed with a specific technique named *partial volume interpolation* (PVI). This technique is necessary to avoid abrupt changes when different image alignments are tried during registration, which could cause unwanted local maxima in the mutual information metric.

The basic joint histogram between two images $\boldsymbol{F}$ and $\boldsymbol{M}$ is a matrix $h(f, m)$, where f and m are intensity bin intervals from images $\boldsymbol{F}$ and $\boldsymbol{M}$ respectively. Each position (f, m) denotes how many times that pair of intensities occurs in corresponding positions in each image. The joint distribution can be estimated by normalizing the histogram as in $p_{FM}(f, m) = h(f, m)/\sum h(f, m)$. Marginal histograms are obtained by summing over a variable as in $p_F(f) = \sum_m p_{FM}(f, m)$.

During registration, intensities from the moving image $\boldsymbol{M}(\mathbf{x})$ are paired with intensities from the fixed image $\boldsymbol{F}(T_\theta \mathbf{x})$, that need to be interpolated. This introduces new values to the histogram in each optimization iteration, causing fluctuation in successive mutual information calculations, which leads to local maxima.

To avoid this, the PVI algorithm chooses to increment the histogram bins that correspond to the nearest neighbors of $\boldsymbol{F}(T_\theta \mathbf{x})$. To guarantee the consistency of the histogram, the sum of all increments should equal unity. Each neighbor n_i would contribute with a certain weight w_i to a linear interpolation of $\boldsymbol{F}(T_\theta \mathbf{x})$, such that $\sum w_i = 1$. Therefore, these weights are conveniently used to increase the corresponding histogram entries. For a 3D image, there are 8 neighbors, so a total of 8 updates are made:

$$\forall i : h(n_i, m) \leftarrow h(n_i, m) + w_i \tag{8}$$

However, even if the PVI technique is used, there can still be unwanted oscillations in the metric. In particular, grid aligning transformations will cause less weight distribution between the bins and alter the MI calculation at that point.

To minimize these oscillations, Pluim [12] recommends a small resampling of the moving image as a preprocessing step. This guarantees different voxel sizes between the images, turning grid aligning transformations improbable.

A variation of MI also suggested by Pluim [13] was used to improve the registration of PET images (e.g., Fig. 1f). It consists in evaluating the alignment of the image gradients for each pair of corresponding points, and using it to weight the original MI. This favors alignments where borders in both images are in corresponding positions. The content of PET images is different than that of anatomical images such as CT or MR, but the head contour is still discernible. Its borders provide therefore good queues for alignment.

B. Fusion criteria

At the end of the registration process, the joint histogram of the aligned images is available. The suggested fusion equations (5)-(7) depend only on the marginal and condition probabilities, which can both be extracted from this histogram. Since obtaining it is a computationally intensive task, we choose to reuse it in the fusion stage by integrating the modules.

Conditional probabilities may be obtained from the histogram $h(f, m)$ as $p(m|f) = h(f, m)/h(f)$ or $p(f|m) = h(f, m)/h(m)$, where $h(f) = \sum_m h(f, m)$ and $h(m) = \sum_f h(f, m)$ are the marginal histograms from images $\boldsymbol{F}$ and $\boldsymbol{M}$ respectively.

With this quantities, the I_1, I_2 and I_3 maps are generated by a straightforward implementation of its equations. Each voxel has a measure associated with its intensity value, so voxels from different images can be compared to specify the fused data set.

We chose two fusion criteria proposed by Bramon *et al.* [7] which, based on our tests, provided the best results:

- Symmetric I_2 fusion:

$$f = \begin{cases} a, & I_2(a, B) > I_2(b, A) \\ b, & \text{otherwise.} \end{cases} \tag{9}$$

- Asymmetric $I_2 I_3$ fusion:

$$f = \begin{cases} a, & I_2(a, B) > I_3(a, B) \\ b, & \text{otherwise.} \end{cases} \tag{10}$$

V. Implementation

We restrain the alignment to a rigid transformation with 6 parameters, consisting of 3 translations in each axis, t_x, t_y and t_z, and 3 rotations, θ_x, θ_y and θ_z. This restriction is reasonable for intrapatient head image registration because of the neurocranium rigid structure.

The joint histogram size is 256×256, and the image intensities are first quantized to these many levels. The histogram construction is a computationally intensive task, therefore some methods to improve the overall efficiency were employed.

Firstly, a parallel approach based on reduction techniques is used to calculate the histogram. Each thread i calculates a histogram $h_i(f, m)$ for a portion $\boldsymbol{M_i}$ of the moving image $\boldsymbol{M}$. Subsequently, the final histogram is given by $h(f, m) = \sum_i h_i(f, m)$.

Secondly, during registration, a multi-resolution strategy is adopted to improve the optimization efficiency. An initial estimate is obtained with less processing by working with less data and simpler transform models. This coarse estimate acquired quickly is fed to later stages, where more processing power can be better aimed at finding fine estimates. A two-level pyramid was created by subsampling the images by an integer factor, usually 4 or 2 for lower resolution images. Also, nearest neighbor interpolation was used, so that no new intensities are introduced. Also, some transformation parameters tend to be smaller due to the patient's positioning,

e.g., rotation in non-axial planes. They are discarded in the first level to speed up optimization. The second level used all available data and optimized the 6 tranform parameters.

The NEWUOA search algorithm developed by Powell [18] was employed to perform the optimization. It is a nonlinear unrestricted gradient free local minimization algorithm. The mutual information gradient calculations using histograms are expensive to compute, so generally a gradient free optimizer is used in literature.

The optimization robustness is influenced by the order in which the parameters are considered. It is recommended [19] to first align the images in the transversal plane by optimizing the in-plane translation and rotation parameters t_x, t_y and θ_z, and after that the out-of-plane parameters.

The fusion measures are calculated using the same joint histogram that is obtained during registration. Our implementation differs from that of Bramon *et al.* [7] because we use a reduced histogram and no preprocessing steps are taken, such as segmentation. This didn't interfere with the results, as demonstrated by the experiments reported in Section VI.

VI. Experiments

We validate our technique through a series of experiments. The registration algorithm implemented in C++ and the fusion stage is currently prototyped in MATLAB. The experimentation platform was a desktop Intel®Core i5 2.8 GHz CPU with 4GB RAM and a NVIDIA GeForce GTS250, 1GB VRAM.

First experiment: The aforementioned registration techniques were implemented in C++ and validated with the Retrospective Image Registration Evaluation (RIRE) [20] project. This database provides axial CT, MR and PET image pairs that were divided in three datasets named I-III detailed in Table I. Datasets I and II contain CT-MR pairs. Dataset II has slightly higher resolution but external artifacts appear in the CT images. Dataset III contains PET-MR pairs, with smaller resolution for the PET images. Gold standards for the rigid registration of these pairs were obtained by clinicians by aligning fiducial markers (later removed from the images). Submitted registration attempts are assessed by comparing the alignment of volumes of interest (VOIs), determined by these fiducial markers.

A separate set of high resolution CT, MR and PET images was provided by our university hospital. It is specified as dataset IV in Table I. They were registered in pairs, and we assessed the results visually to guarantee they could be used in the next experiments regarding fusion and visualization. The MR study was sagittally oriented, while CT and PET images were oriented axially.

Second experiment: After obtaining the spatial mapping between image pairs in the last experiment, we applied the fusion criteria based on specific information to them. Currently we have prototyped the fusion measures in MATLAB to test them out in different combinations for evaluation purposes. We simulate the construction of the joint histogram generated during the registration algorithm and calculate the measures from it according to Section IV-B.

TABLE I
Images used throughout the experiments.

Dataset	Modalities	Dimensions	Voxel size (mm)	Pairs
I	CT	$512^2 \times$ (27–34)	$0.65^2 \times 4.0$	41
	MR T1/T2/PD	$256^2 \times$ (20–26)	$1.25^2 \times 4.0$	
II	CT	$512^2 \times$ (40–49)	$(0.40–0.45)^2 \times 3.0$	21
	MR T1/T2/PD	$256^2 \times$ (51–52)	$(0.78–0.86)^2 \times 3.0$	
III	PET	$128^2 \times 15$	$2.59^2 \times 8.0$	35
	MR	$256^2 \times$ (24–26)	$1.25^2 \times 4.0$	
IV	CT	$512^2 \times 110$	$0.49^2 \times 1.5$	3
	MR T1	$240^2 \times 180$	$1.0^2 \times 1.0$	
	PET	$256^2 \times 110$	$1.33^2 \times 1.5$	

TABLE II
Registration results for the RIRE database (errors in mm).

	errors			
Dataset	mean	maximum	median	failures
I	1.721	5.179	1.469	0
II	1.923	3.480	1.879	0
III	2.644	6.827	2.399	1

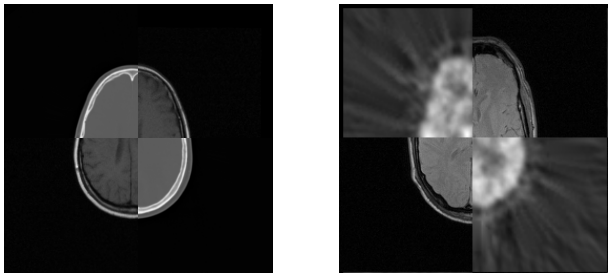
The same images from the previous experiment were used. Since there is no meaningful quantitative measure of the fusion results, we assess it visually by combining the selected voxels from each image in different color channels.

VII. Results and Discussion

Quantitative tests based on gold standards were used to validate the registration step, while fusion and visualization were evaluated by visual assessment, as detailed below.

A. Quality

The validation of the registration algorithm with the RIRE database is summarized in Table II, where the errors in the alignment of VOIs are reported. The accuracy of the results is directly related to the images' resolution. We consider a registration successful if the average distance of corresponding VOIs is smaller than the greater slice thickness. The average error throughout the database is acceptable, and there was only one failure case for PET registration. Fig. 1 and Fig. 3 show examples of combined images in a checker board pattern after registration from each dataset.



(a) (b)

Fig. 3. Registered images on a checkerboard pattern from datasets a) I and b) III.

Fusion results are illustrated in Fig. 1. Fig. 1d) is an asymmetric I_2I_3 fusion between CT and MR images from dataset II with CT as reference. It can be seen that bone structures are mostly selected from CT. This is desirable since they have better contrast in this technology. Remaining structures correspond to soft tissue areas and are selected from MR. Fig. 1h) is an asymmetric I_2I_3 fusion between CT and PET images from dataset IV with PET as reference. Similarly, bone structures are selected from CT, while brain activation areas from PET. This would provide a clinician with an image that shows both functional information and anatomical cues.

B. Performance

The registration of images from dataset I takes under 1 minute, and under 2 minutes for dataset II using the hardware detailed in VI. This is consistent with some results in literature [21] but can still be improved.

The fusion stage was implemented as a prototype in MATLAB, and, after the joint histogram was built, each measure took in the order of hundreds of milliseconds to be computed. The actual fusion criteria took in the order of tens of milliseconds. An implementation in C++ would surely deliver real time responsiveness.

C. Limitation

The registration accuracy is dependent on the volumes' spatial resolution, and there are also efficiency figures that can be improved. The fusion criteria and tuning parameter still must be chosen interactively. While this can be viewed as an additional step moving away from the task automation, it can be considered an opportunity for a trained operator to contribute to the final result and tailor it to his needs.

VIII. Conclusion

In this paper, we presented a unified framework for aligning and combining multimodal images with mutual information based measures. Prototyping with MATLAB has been successfully conducted. Both registration's and fusion's quality can be considered satisfactory. Currently, we are porting all algorithms to C++ and, as reported in Section V, performance is an important issue.

Our experiments have been performed in view of computational resources and non-expert subjectiveness. Nevertheless, to assess its practical usefulness the outcome must be validated clinically by the experts in the medical field. We plan to realize this validation when we conclude the code porting and the integration of an information-preserved volume render that is under development.

Acknowledgment

The authors would like to thank their colleagues at their university hospital, in particular Dr. Clarissa L. Yasuda, for providing medical images. This work has been supported by CNPq and Fapesp (grant 2011/2351-0).

References

[1] D. Valentino, J. Mazziotta, and H. Huang, "Volume rendering of multimodal images: application to mri and pet imaging of the human brain," *Medical Imaging, IEEE Transactions on*, vol. 10, no. 4, pp. 554 –562, Dec. 1991.

[2] R. Frank, H. Damasio, and T. Grabowski, "Brainvox: An interactive, multimodal visualization and analysis system for neuroanatomical imaging," *ni*, vol. 5, pp. 13–30, 1997.

[3] T. Ropinski *et al.*, "Surface glyphs for visualizing multimodal volume data," in *In Proceedings of the 12th International Fall Workshop on Vision, Modeling, and Visualization (VMV07*, 2007, pp. 3–12.

[4] C. Rieder, M. Schwier, H. K. Hahn, and H.-O. Peitgen, "High-Quality Multimodal Volume Visualization of Intracerebral Pathological Tissue." Eurographics Workshop on Visual Computing for Biomedicine, 2008, pp. 167–176.

[5] F. Maes, D. Vandermeulen, and P. Suetens, "Medical image registration using mutual information," *Proceedings of the IEEE*, vol. 91, no. 10, pp. 1699–1722, 2003.

[6] M. Haidacher, S. Bruckner, A. Kanitsar, and M. E. Groller, "Information-based transfer functions for multimodal visualization," in *VCBM*, W. N. C.P Botha, G. Kindlmann and B. Preim, Eds. Eurographics Association, Oct. 2008, pp. 101–108.

[7] R. Bramon *et al.*, "Multimodal data fusion based on mutual information," *IEEE Transactions on Visualization and Computer Graphics*, vol. 99, no. PrePrints, 2011.

[8] E. Gulsoy, J. Simmons, and M. Degraef, "Application of joint histogram and mutual information to registration and data fusion problems in serial sectioning microstructure studies," *Scripta Materialia*, vol. 60, no. 6, pp. 381–384, 2009.

[9] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Physics in Medicine and Biology*, vol. 46, no. 3, pp. R1–R45, Mar. 2001.

[10] A. Collignon *et al.*, *Automated multi-modality image registration based on information theory*. Kluwer Academic Publishers, 1995, vol. 3, no. 6, pp. 263–274.

[11] P. Viola and W. M. Wells, III, "Alignment by maximization of mutual information," in *International Journal of Computer Vision*, 1995, pp. 16–23.

[12] J. Pluim, J. Maintz, and M. Viergever, "Interpolation artefacts in mutual information based image registration," *Computer Vision and Image Understanding*, vol. 77, no. 2, pp. 211–232, 2000.

[13] J. P. Pluim, J. B. Maintz, and M. A. Viergever, "Image registration by maximization of combined mutual information and gradient information." *IEEE transactions on medical imaging*, vol. 19, no. 8, pp. 809–14, Aug. 2000.

[14] Z. Wang *et al.*, "A comparative analysis of image fusion methods," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1391–1402, Jun. 2005.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 99th ed. Wiley-Interscience, Aug. 1991.

[16] M. R. DeWeese and M. Meister, "How to measure the information gained from one symbol." *Network: Computation in Neural Systems*, vol. 10, no. 4, pp. 325–340, Nov. 1999.

[17] D. A. Butts, "How much information is associated with a particular stimulus?" *Network (Bristol, England)*, vol. 14, no. 2, pp. 177–187, May 2003.

[18] M. J. D. Powell, "The NEWUOA software for unconstrained optimization without derivatives," *LargeScale Nonlinear Optimization*, vol. 83, pp. 255–297, 2006.

[19] F. Maes *et al.*, "Multimodality image registration by maximization of mutual information," *IEEE transactions on Medical Imaging*, vol. 16, pp. 187–198, 1997.

[20] J. West *et al.*, "Comparison and evaluation of retrospective intermodality brain image registration techniques," *Journal of Computer Assisted Tomography*, vol. 21, pp. 554–566, 1998.

[21] R. Shams, P. Sadeghi, R. A. Kennedy, and R. Hartley, "Parallel computation of mutual information on the GPU with application to real-time registration of 3D medical images," *Computer Methods and Programs in Biomedicine*, vol. 99, no. 2, pp. 133–146, Aug. 2010.