A VARIATIONAL APPROACH FOR OPTIMIZING QUADRATIC MUTUAL INFORMATION FOR MEDICAL IMAGE REGISTRATION

Abhishek Singh[†*], *Ying Zhu*[‡] *and Christophe Chefd'hotel*[‡]

[†]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, IL, USA
[‡]Siemens Corporate Research, Princeton, NJ, USA
abhishek_singh@ieee.org, {yingzhu, christophe.chefdhotel}@siemens.com

ABSTRACT

This paper explores the use of quadratic mutual information as a similarity criterion for dense, non-rigid registration of medical images. Quadratic mutual information between two random variables has been recently proposed as Euclidean distance between the joint density and the product of the marginals. It has been shown to have a smooth sample estimator, that can be computed without having to use numerical approximation techniques for computing the integral over densities. In this paper, we derive Euler-Lagrange equations for optimizing quadratic mutual information in a variational framework. We then obtain a dense deformation field for registering 3D tomography images. Our results demonstrate the applicability of this criterion for such a task, and yield ground for further analysis and research.

***Index Terms*—** Information theoretic learning, quadratic mutual information, variational calculus, medical image registration

1. INTRODUCTION

Matching or aligning images is a fundamental problem, and a very important issue in medical image processing communities. The need for alignment or registration algorithms arises, for instance, while comparing medical images taken at different instances in time, that may get misaligned due to respiratory motion, cardiac motion etc. Registration is also required when images from multiple acquisition modalities must be realigned for better fusion of complementary information [1].

The general framework for registration of images involves learning the parameters of the transformation between the two images by optimizing an intensity based similarity measure between the first image and the transformed or warped second image. The class of transformations can be low-dimensional, parametric transforms such as affine or rigid, or can be more complex parametric models [2]. Variational methods are quite popular as they can recover a dense displacement field by optimizing a similarity measure over a suitable functional space in a variational framework, allowing it to handle complex, non-rigid transformations. Chefd'hotel and Hermosillo [1, 3] formulate a variety of similarity measures in such a variational framework.

Mutual information (MI) or relative entropy has been a widely used cost function for intensity based registration of medical images, since it was first proposed in [4]. Mutual information is a measure of dispersion of the joint (2-D) density of the intensities of corresponding voxel pairs in the images. Mutual information is maximized and dispersion minimized if the two images are geometrically aligned.

The definition of mutual information that is based on Shannon's definition of entropy essentially computes the Kullback-Leibler divergence between the joint density and the product of the marginal densities of the two random variables:

$$\mathcal{I} = \int p(i_1, i_2) \log \left(\frac{p(i_1, i_2)}{p(i_1)p(i_2)} \right) d\mathbf{i} \tag{1}$$

This definition of mutual information requires integrating over the probability densities. Unless the densities are assumed to be in simple analytical forms, this integration needs to be approximated using numerical techniques, which is off-putting for practical applications.

It has been shown by Principe et al. [5] that using alternative measures of distance/divergence between densities can allow for analytical and exact computation of this integral.

Quadratic Mutual Information - Euclidean Distance (QMI-ED) between two PDFs is defined by Principe et al. [5] as the Euclidean distance between the joint and the product of the marginal densities:

$$\mathcal{I}_{ED} = \int \left(p(i_1, i_2) - p(i_1)p(i_2)\right)^2 d\mathbf{i} \tag{2}$$

We use the above definition for computing mutual information between two CT images, and show that the integral over the image intensities can be analytically computed, without having to use numerical methods for approximating it.

For registering two images, we obtain the warping function between them by optimizing QMI-ED using a variational approach, along the lines of [1]. We derive the Euler-Lagrange equations for QMI-ED, and solve them using a gradient descent strategy, obtaining a dense, non-linear displacement field that aligns the two images.

In the next section, we briefly describe how the image registration problem is framed using variational calculus [6, 1]. In Section 3, we show how QMI-ED can be used in this variational setting, and derive the Euler-Lagrange equations for it. Section 4 discusses some implementation details and Section 5 presents some results.

2. VARIATIONAL FRAMEWORK FOR REGISTRATION

Consider two n-dimensional ($n = 2, 3$) images I_1 and I_2, defined over a spatial region Ω (a bounded region of $\Re^n$). To register the two images, we look for a function $\mathbf{h} : \Omega \rightarrow \Re^n$ assigning to each point

*Work was done while at Siemens Corporate Research, Princeton, NJ, USA

$\mathbf{x}$ in Ω a displacement vector $\mathbf{h}(\mathbf{x}) \in \Re^n$, such that it minimizes an energy functional of the form

$$\mathcal{J}(\mathbf{h}) = \mathcal{I}(\mathbf{h}) + \alpha R(\mathbf{h}). \tag{3}$$

The term $\mathcal{I}(\mathbf{h})$ measures the dissimilarity between the first image $I_1(\mathbf{x})$ and the warped second image $I_2(\mathbf{x} + \mathbf{h}(\mathbf{x}))$. The term $R(\mathbf{h})$ is a regularization term designed to penalize fast variations of the function $\mathbf{h}$. The first variation [6] of $\mathcal{J}$ at $\mathbf{h}$ in the direction of $\mathbf{k}$ (Gateaux derivative) is defined as,

$$\delta_{\mathbf{k}}\mathcal{J}(\mathbf{h}) = \left.\frac{\partial \mathcal{J}(\mathbf{h}+\epsilon\mathbf{k})}{\partial \epsilon}\right|_{\epsilon=0} \tag{4}$$

The gradient $\nabla\mathcal{J}(\mathbf{h})$ is defined by requiring the equality

$$\delta_{\mathbf{k}}\mathcal{J}(\mathbf{h}) = \langle \nabla\mathcal{J}(\mathbf{h}), \mathbf{k}\rangle \tag{5}$$

to hold for every $\mathbf{k}$. If a minimizer $\mathbf{h}$ of $\mathcal{J}$ exists, then the set of equations $\delta_{\mathbf{k}}\mathcal{J}(\mathbf{h}) = 0$ must hold for every $\mathbf{k}$, which implies $\nabla\mathcal{J}(\mathbf{h}) = 0$. These are called the Euler-Lagrange equations [6] associated with the energy functional $\mathcal{J}$. Given a particular form of the energy functional, its functional derivative can be obtained by first computing its Gateaux derivative, and then expressing it in the form of the inner product of Eqn. 5.

Since an exact analytic solution for the Euler-Lagrange equations is rarely possible for commonly used energy functionals, the search for the minimizer is generally done using a gradient descent strategy, where gradient can be expressed as,

$$\nabla\mathcal{J}(\mathbf{h}) = \nabla\mathcal{I}(\mathbf{h}) + \alpha\nabla R(\mathbf{h}) \tag{6}$$

The energy functional contains a regularization term that a function of the Jacobian $D\mathbf{h}$ of the form,

$$R(\mathbf{h}) = \int_\Omega \phi(D\mathbf{h}(\mathbf{x}))d\mathbf{x} \tag{7}$$

Hermosillo et al. [1] and Alvarez et al. [7] use a regularization function $\phi(.)$ based on the Nagel and Enkelmann operator [8], which has an efficient estimation scheme as described in [1, 8]. We use the same in our implementation.

3. VARIATIONAL GRADIENT FOR QUADRATIC MUTUAL INFORMATION

We focus our attention in this paper on the first term of the energy gradient (first term in the RHS of Eqn. 6). In particular, we describe how this gradient can be computed for the quadratic mutual information criteria, in the variational framework as discussed above.

Let $i_1 = I_1(\mathbf{x})$, $i_2 = I_2(\mathbf{x} + \mathbf{h}(\mathbf{x}))$, $\mathbf{i} = (i_1, i_2)$ and $\mathbf{I_h}(\mathbf{x}) = [I_1(\mathbf{x}), I_2(\mathbf{x}+\mathbf{h}(\mathbf{x}))]$. QMI-ED between $I_1(\mathbf{x})$ and the warped image $I_2(\mathbf{x} + \mathbf{h}(\mathbf{x}))$ is defined as,

$$\begin{aligned}\mathcal{I}_{ED}(\mathbf{h}) &= \int_{\Re^2} \left(p(\mathbf{i},\mathbf{h}) - p(i_1)p(i_2,\mathbf{h})\right)^2 d\mathbf{i} && (8)\\ &= \int_{\Re^2} p^2(\mathbf{i},\mathbf{h})d\mathbf{i} + \int_{\Re^2} p^2(i_1)p^2(i_2,\mathbf{h})d\mathbf{i} \\ &\quad -2\int_{\Re^2} p(\mathbf{i},\mathbf{h})p(i_1)p(i_2,\mathbf{h})d\mathbf{i} && (9)\end{aligned}$$

Let us denote the three terms in the above expression as $\mathcal{I}_{ED1}(\mathbf{h})$, $\mathcal{I}_{ED2}(\mathbf{h})$ and $\mathcal{I}_{ED3}(\mathbf{h})$, such that $\mathcal{I}_{ED}(\mathbf{h}) = \mathcal{I}_{ED1}(\mathbf{h}) + \mathcal{I}_{ED2}(\mathbf{h}) - \mathcal{I}_{ED3}(\mathbf{h})$.

The estimates of probability densities are computed using Parzen windowing technique as follows,

$$\begin{aligned}p(\mathbf{i},\mathbf{h}) &= \frac{1}{|\Omega|}\int_\Omega G_\Sigma(\mathbf{I_h}(\mathbf{x}) - \mathbf{i})d\mathbf{x} && (10)\\ p(i_1) &= \frac{1}{|\Omega|}\int_\Omega G_\sigma(I_1(\mathbf{x}) - i_1)d\mathbf{x} && (11)\\ p(i_2,\mathbf{h}) &= \frac{1}{|\Omega|}\int_\Omega G_\sigma(I_2(\mathbf{x}+\mathbf{h}(\mathbf{x})) - i_2)d\mathbf{x} && (12)\end{aligned}$$

where $G_\sigma(.)$ denotes a Gaussian function with width parameter σ.

We now substitute these non-parametric density estimates back into each of the three terms of $\mathcal{I}_{ED}(\mathbf{h})$. The first term, $\mathcal{I}_{ED1}(\mathbf{h})$, now becomes,

$$\begin{aligned}\mathcal{I}_{ED1}(\mathbf{h}) &= \frac{1}{|\Omega|^2}\int_\Omega\int_\Omega\int_{\Re^2} G_\Sigma(\mathbf{I_h}(\mathbf{x}) - \mathbf{i})G_\Sigma(\mathbf{I_h}(\mathbf{y}) - \mathbf{i})d\mathbf{i}\, d\mathbf{x}d\mathbf{y} \\ &= \frac{1}{|\Omega|^2}\int_\Omega\int_\Omega G_{2\Sigma}(\mathbf{I_h}(\mathbf{x}) - \mathbf{I_h}(\mathbf{y}))d\mathbf{x}d\mathbf{y} && (13)\end{aligned}$$

The above expression is obtained by using the convolution theorem of Gaussians, which states that the convolution of two Gaussian functions yields another Gaussian with a scaled width parameter. This property of the Gaussian kernel helps us obtain an exact analytic solution for the integral over the image intensities, which is not possible in case of using mutual information as defined using Shannon's entropy [5].

The same property can be applied to $\mathcal{I}_{ED2}(\mathbf{h})$ and $\mathcal{I}_{ED3}(\mathbf{h})$ as well, to yield,

$$\begin{aligned}\mathcal{I}_{ED2}(\mathbf{h}) &= \frac{1}{|\Omega|^4}\int_\Omega\int_\Omega G_{2\sigma}(I_1(\mathbf{x}) - I_1(\mathbf{y}))d\mathbf{x}d\mathbf{y} && (14)\\ &\quad\times\int_\Omega\int_\Omega G_{2\sigma}(I_2(\mathbf{x}+\mathbf{h}(\mathbf{x})) - I_2(\mathbf{y}+\mathbf{h}(\mathbf{y})))d\mathbf{x}d\mathbf{y}\end{aligned}$$

$$\begin{aligned}\mathcal{I}_{ED3}(\mathbf{h}) &= \frac{2}{|\Omega|^3}\int_\Omega\int_\Omega\int_\Omega G_{2\sigma}(I_1(\mathbf{x}) - I_1(\mathbf{y})) && (15)\\ &\quad\times G_{2\sigma}(I_2(\mathbf{x}+\mathbf{h}(\mathbf{x})) - I_2(\mathbf{z}+\mathbf{h}(\mathbf{z})))\, d\mathbf{x}d\mathbf{y}d\mathbf{z}\end{aligned}$$

We now compute the Gateaux derivative for $\mathcal{I}_{ED1}$, $\mathcal{I}_{ED2}$ and $\mathcal{I}_{ED3}$. The first variation of $\mathcal{I}_{ED1}$ at $\mathbf{h}$ in the direction of $\mathbf{k}$ is defined as,

$$\begin{aligned}&\delta_{\mathbf{k}}\mathcal{I}_{ED1}(\mathbf{h}) = \left.\frac{\partial \mathcal{I}_{ED1}(\mathbf{h}+\epsilon\mathbf{k})}{\partial\epsilon}\right|_{\epsilon=0}\\ &= \frac{-1}{\sigma^2|\Omega|^2}\int_\Omega\int_\Omega G_{2\Sigma}\left(\mathbf{I_h}(\mathbf{x}) - \mathbf{I_h}(\mathbf{y})\right)\cdot[I_2(\mathbf{x}+\mathbf{h}(\mathbf{x})) - I_2(\mathbf{y}+\mathbf{h}(\mathbf{y}))]\\ &\quad\cdot[\nabla I_2(\mathbf{x}+\mathbf{h}(\mathbf{x}))\mathbf{k}(\mathbf{x}) - \nabla I_2(\mathbf{y}+\mathbf{h}(\mathbf{y}))\mathbf{k}(\mathbf{y})]\ d\mathbf{x}d\mathbf{y}\\ &= \frac{-1}{\sigma^2|\Omega|^2}\int_\Omega\int_\Omega G_{2\Sigma}\left(\mathbf{I_h}(\mathbf{x}) - \mathbf{I_h}(\mathbf{y})\right)\cdot[I_2(\mathbf{x}+\mathbf{h}(\mathbf{x})) - I_2(\mathbf{y}+\mathbf{h}(\mathbf{y}))]\\ &\quad\cdot\nabla I_2(\mathbf{x}+\mathbf{h}(\mathbf{x}))\mathbf{k}(\mathbf{x})\ d\mathbf{x}d\mathbf{y}\\ &\quad+\frac{1}{\sigma^2|\Omega|^2}\int_\Omega\int_\Omega G_{2\Sigma}\left(\mathbf{I_h}(\mathbf{x}) - \mathbf{I_h}(\mathbf{y})\right)\cdot[I_2(\mathbf{x}+\mathbf{h}(\mathbf{x})) - I_2(\mathbf{y}+\mathbf{h}(\mathbf{y}))]\\ &\quad\cdot\nabla I_2(\mathbf{y}+\mathbf{h}(\mathbf{y}))\mathbf{k}(\mathbf{y})\ d\mathbf{x}d\mathbf{y} && (16)\end{aligned}$$

By interchanging the $\mathbf{x}$ and $\mathbf{y}$ variables in the second term of the RHS above, and by using the fact that the Gaussian kernel is symmetric, we can write $\delta_{\mathbf{k}}\mathcal{I}_{ED1}(\mathbf{h})$ in an inner product form as,

$$\delta_{\mathbf{k}}\mathcal{I}_{ED1}(\mathbf{h}) = \int_\Omega \mathbf{k}(\mathbf{x})F_1(\mathbf{x})d\mathbf{x} \tag{17}$$

where,

$$F_1(\mathbf{x}) = \frac{-2}{\sigma^2|\Omega|^2} \cdot \nabla I_2(\mathbf{x}+\mathbf{h}(\mathbf{x})) \int_\Omega G_{2\Sigma}\left(\mathbf{I_h}(\mathbf{x}) - \mathbf{I_h}(\mathbf{y})\right) \cdot \left[I_2(\mathbf{x}+\mathbf{h}(\mathbf{x})) - I_2(\mathbf{y}+\mathbf{h}(\mathbf{y}))\right] d\mathbf{y} \tag{18}$$

Using similar algebraic manipulations for computing the first variation of $\mathcal{I}_{ED2}$, we get,

$$\delta_{\mathbf{k}}\mathcal{I}_{ED2}(\mathbf{h}) = \left.\frac{\partial \mathcal{I}_{ED2}(\mathbf{h}+\epsilon\mathbf{k})}{\partial \epsilon}\right|_{\epsilon=0} = \int_\Omega \mathbf{k}(\mathbf{x}) F_2(\mathbf{x}) d\mathbf{x} \tag{19}$$

where

$$F_2(\mathbf{x}) = C_2 \cdot \frac{-2}{\sigma^2|\Omega|^2} \cdot \nabla I_2(\mathbf{x}+\mathbf{h}(\mathbf{x})) \times \int_\Omega G_{2\sigma}(I_2(\mathbf{x}+\mathbf{h}(\mathbf{x})) - I_2(\mathbf{y}+\mathbf{h}(\mathbf{y}))) \cdot \left[I_2(\mathbf{x}+\mathbf{h}(\mathbf{x})) - I_2(\mathbf{y}+\mathbf{h}(\mathbf{y}))\right] d\mathbf{y} \tag{20}$$

and

$$C_2 = \frac{1}{|\Omega|^2} \int_\Omega \int_\Omega G_{2\sigma}(I_1(\mathbf{x}) - I_1(\mathbf{y})) d\mathbf{x} d\mathbf{y} \tag{21}$$

Similarly, after some algebraic manipulations (not shown here due to space constraints), we can express the Gateaux derivative of $\mathcal{I}_{ED3}(\mathbf{h})$ as the following inner product:

$$\delta_{\mathbf{k}}\mathcal{I}_{ED3}(\mathbf{h}) = \left.\frac{\partial \mathcal{I}_{ED3}(\mathbf{h}+\epsilon\mathbf{k})}{\partial \epsilon}\right|_{\epsilon=0} = \int_\Omega \mathbf{k}(\mathbf{x}) F_3(\mathbf{x}) d\mathbf{x} \tag{22}$$

where

$$F_3(\mathbf{x}) = \frac{-2}{\sigma^2|\Omega|^3} \cdot \nabla I_2(\mathbf{x}+\mathbf{h}(\mathbf{x})) \times \int_\Omega [A(\mathbf{x}) + A(\mathbf{z})] \cdot G_{2\sigma}(I_2(\mathbf{x}+\mathbf{h}(\mathbf{x})) - I_2(\mathbf{z}+\mathbf{h}(\mathbf{z}))) \cdot \left[I_2(\mathbf{x}+\mathbf{h}(\mathbf{x})) - I_2(\mathbf{z}+\mathbf{h}(\mathbf{z}))\right] d\mathbf{z} \tag{23}$$

and

$$A(\mathbf{x}) = \int_\Omega G_{2\sigma}(I_1(\mathbf{x}) - I_1(\mathbf{y})) d\mathbf{y}. \tag{24}$$

Therefore, using Eqns. 17, 19 and 22, the variational gradient $\nabla\mathcal{I}_{ED}(\mathbf{h})$ is given by,

$$\nabla\mathcal{I}_{ED}(\mathbf{h})(\mathbf{x}) = F_1(\mathbf{x}) + F_2(\mathbf{x}) - F_3(\mathbf{x}), \tag{25}$$

where $F_1(\mathbf{x})$, $F_2(\mathbf{x})$ and $F_3(\mathbf{x})$ are defined by Eqns. 18, 20 and 23.

4. IMPLEMENTATION DETAILS

To summarize, the proposed algorithm involves the following steps to compute the optimal warping function $\mathbf{h}(\mathbf{x})$, to register the images $I_1(\mathbf{x})$ and $I_2(\mathbf{x}+\mathbf{h}(\mathbf{x}))$: For each location $\mathbf{x}$, we first evaluate the gradient components $F_1(\mathbf{x})$, $F_2(\mathbf{x})$ and $F_3(\mathbf{x})$ using Eqns. 18, 20 and 23. We then use Eqn. 25 to obtain the gradient $\nabla\mathcal{I}_{ED}(\mathbf{h})(\mathbf{x})$ of the QMI-ED functional. We then update the warping function $\mathbf{h}(\mathbf{x})$ using gradient descent steps as follows:

$$\begin{aligned} \mathbf{h}^+(\mathbf{x}) &= \mathbf{h}(\mathbf{x}) - \mu\nabla\mathcal{J}(\mathbf{h})(\mathbf{x}) \\ &= \mathbf{h}(\mathbf{x}) - \mu\left[\nabla\mathcal{I}_{ED}(\mathbf{h})(\mathbf{x}) + \alpha\nabla R(\mathbf{h})(\mathbf{x})\right] \end{aligned} \tag{26}$$

where the regularization functional $R(\mathbf{h})$ is as described in Section 2.

Each of $F_1(\mathbf{x})$, $F_2(\mathbf{x})$ and $F_3(\mathbf{x})$ are essentially weighted sums of Gaussian evaluations. In our implementation, we leverage on the computational speed up provided by the Fast Gauss Transform [9] as suggested in [5], for computing these expressions.

Our algorithm requires selecting a kernel size parameter σ for computing the Gaussian evaluations. A detailed study of the effects of the kernel size can be found in [10, 5]. For our registration application, the kernel size was set to unity, after normalizing the image intensities to have unit variance.

To avoid local optima and for faster convergence, a coarse-to-fine approach is used for the optimization procedure, wherein warping function is first computed on a subsampled (coarse) version of the original volume, and the estimates are refined by using successively higher resolutions. Three levels of such a pyramid approach were used with 16, 8 and 2 iterations in each level.

5. EXPERIMENTS AND RESULTS

Before applying the proposed variational framework for registering images, we first run some simple experiments to study the response surface of the QMI-ED cost function, using synthetic data. We compute QMI-ED between a *sine* function and a translated/displaced version of itself, under additive noise, and plot the cost function value for different values of the displacement parameter. Fig. 1 compares the response surfaces obtained with QMI-ED and the conventional mutual information (MI) of Eqn. 1, for different additive noise levels. Clearly, QMI-ED is a much smoother cost function to optimize under noisy conditions, and better suited for gradient based optimization as compared to conventional MI.

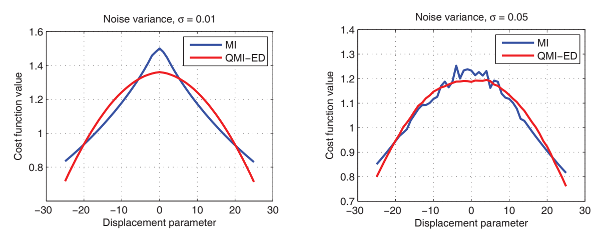


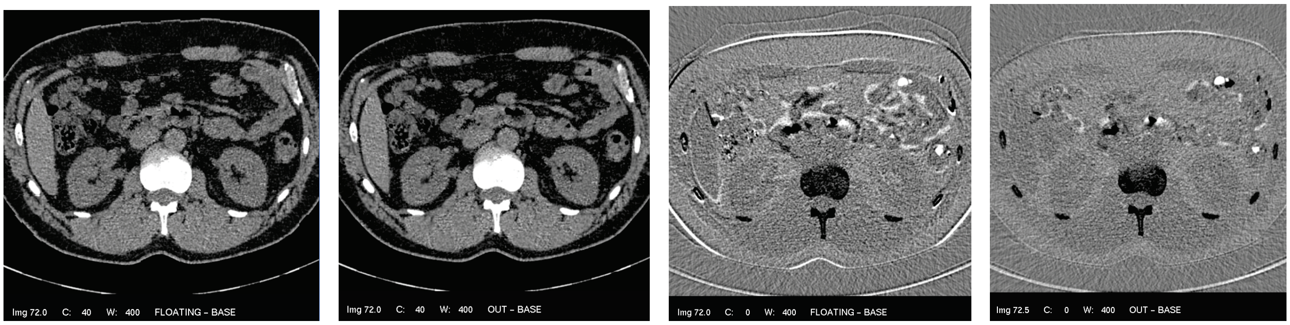
Fig. 1. A comparison of the response surfaces obtained using QMI-ED and conventional mutual information (MI), while registering displaced *sin* functions under different levels of additive noise.

We now test our variational algorithm for registering dual energy computerized tomography (CT) images of human abdominal region. We test on two image pairs. Each unregistered pair consists of a *base* image, and a *floating* image which needs to be warped to align with the *base* image. Each image is a 3D volume composed of 512^3 voxels. The unregistered images have been obtained using different CT energy levels, and are misaligned due to patient movements etc.

Figs. 2 and 3 show our registration results. Fig. 2(a) shows the unregistered volumes (*base* and *floating*) in a checkerboard view (alternate square patches from the two volumes). Misalignment can be seen upon zooming in along the edges. Fig. 2(b) shows the checkerboard view of the registered volumes (*base* and *output*). The misalignments have been corrected.

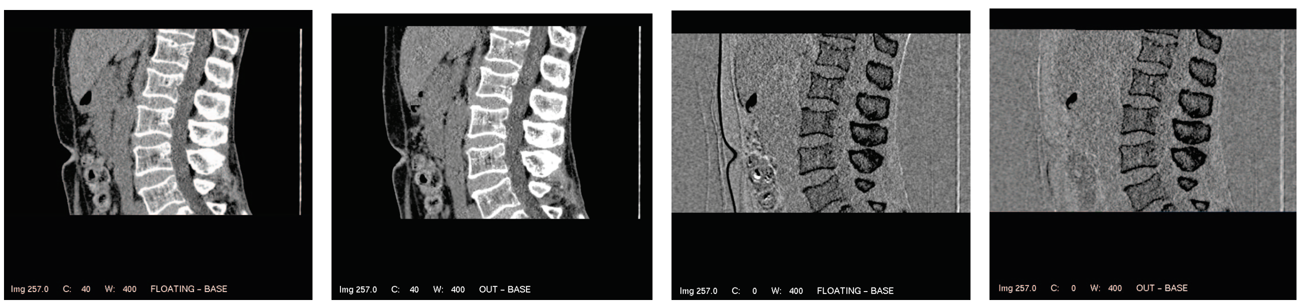
Fig. 2(c) shows the image of voxel-wise intensity difference between the unregistered volumes. Fig. 2(d) shows the difference image computed from the registered volumes. The registered images produce a smoother difference image.

Fig. 3 shows results on the second image pair, along the saggital view of the volumes. Again, the misalignment that is seen along



(a) Floating-Base checkerboard (b) Output-Base checkerboard (c) Floating-Base difference (d) Output-Base difference

Fig. 2. Test Image 1 (Axial view) - (a) Checkerboard image of the two unregistered volumes, obtained by interleaving alternate square patches extracted from the floating and base volumes. (b) Checkerboard image of the registered volumes. (c) Voxel-wise intensity difference between the unregistered volumes. (d) Difference image obtained after registration.



(a) Floating-Base checkerboard (b) Output-Base checkerboard (c) Floating-Base difference (d) Output-Base difference

Fig. 3. Test Image 2 (Saggital view) - (a) Checkerboard image of the unregistered volumes. (b) Checkerboard image of the registered volumes. (c) Voxel-wise intensity difference between the unregistered volumes. (d) Difference image obtained after registration.

body contours in Fig. 3(a) is corrected in Fig. 3(b). The difference image obtained from the registered volumes (Fig. 3(d)) is also much smoother than that obtained with the unregistered volumes (Fig. 3(c)), indicating the effectiveness of the registration procedure.

6. CONCLUSION

In this paper, we have presented a variational approach for optimizing quadratic mutual information, and have applied it for dense, nonrigid registration of CT volumes. QMI-ED has the appeal that it does not require numerically approximating integrals over probability densities, and is a smoother function for optimization as compared to conventional MI, particularly in presense of noise. It has been applied in a variety of signal processing and machine learning applications such as robust ICA and non-linear feature extraction [5] etc. This paper extends its applicability to image registration by presenting a variational scheme for optimization. We have obtained promising results which yield fertile ground for further study and analysis.

7. REFERENCES

[1] Gerardo Hermosillo, Christophe Chefd'Hotel, and Olivier Faugeras, "Variational methods for multimodal image matching," *International Journal of Computer Vision*, vol. 50, pp. 329–343, 2002.

[2] D Rueckert, C Hayes, S Colin, P Summers, M Leach, and D Hawkes, "Non-rigid registration of breast mr images using mutual information," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 1998.

[3] C. Chefd'hotel, G. Hermosillo, and O. Faugeras, "Flows of diffeomorphisms for multimodal image registration," in *IEEE International Symposium on Biomedical Imaging*, 2002, pp. 753 – 756.

[4] P Viola and W.M. Wells, "Alignment by maximization of mutual information," *International Journal of Computer Vision*, vol. 24, pp. 137–154, 1997.

[5] Jose C. Principe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*, Springer, 2010.

[6] R Courant, *Calculus of Variations*, New York, 1946.

[7] Luis Alvarez, Joachim Weickert, and Javier Sánchez, "Reliable estimation of dense optical flow fields with large displacements," *International Journal of Computer Vision*, vol. 39, pp. 41–56, August 2000.

[8] H H Nagel and W Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, pp. 565–593, September 1986.

[9] C Yang, R Duraiswami, N Gumerov, and L Davis, "Improved fast gauss transform and efficient kernel density estimation," in *In International Conference on Computer Vision*, 2003, pp. 464–471.

[10] Abhishek Singh and Jose Principe, "Information theoretic learning with adaptive kernels," *Signal Processing*, vol. 91, no. 2, pp. 203–2013, 2010.