Learning-Based 2D/3D Rigid Registration Using Jensen-Shannon Divergence for Image-Guided Surgery

Rui Liao, Christoph Guetter, Chenyang Xu, Yiyong Sun,
Ali Khamene, and Frank Sauer

Imaging and Visualization Department, Siemens Corporate Research, iPrinceton, NJ, USA

Abstract. Registration of 3D volumetric data to 2D X-ray images has many applications in image-guided surgery, varying from verification of patient position to working projection searching. In this work, we propose a learning-based method that incorporates the prior information on the expected joint intensity histogram for robust real-time 2D/3D registration. Jensen-Shannon divergence (JSD) is used to quantify the statistical (dis)similarity between the observed and expected joint histograms, and is shown to be superior to Kullback-Leibler divergence (KLD) in its symmetry, being theoretically upper-bounded, and well-defined with histogram non-continuity. A nonlinear histogram mapping technique is proposed to handle the intensity difference between the observed data and the training data so that the learned prior can be used for registration of a wide range of data subject to intensity variations. We applied the proposed method on synthetic, phantom and clinical data. Experimental results demonstrated that a combination of the prior knowledge and the low-level similarity measure between the images being registered led to a more robust and accurate registration in comparison with the cases where either of the two factors was used alone as the driving force for registration.

1 Introduction

Techniques for 2D/3D registration can be divided into two general groups: feature-based methods and intensity-based methods. Feature-based methods [1-2] register salient features that have been segmented automatically or semi-manually. While this approach exhibits fast execution time and high robustness in the face of large misregistration, the final accuracy relies on the accuracy of the preprocessing step, i.e. segmentation. In comparison, intensity-based methods [3-4] use the entire image information for registration, and have been shown to give substantially more accurate and reliable results than their feature-based counterparts.

In recent years, learning-based methods have been suggested for general medical registration to impose prior knowledge to achieve more robust and reliable registration. In [5] the log likelihood of the joint intensity distribution of the observed images was maximized with respect to the expected joint intensity distribution. In [6] it was shown empirically that using the Kullback-Leibler divergence (KLD) was superior to the log likelihood measurement. In addition, as a section of experiments in [6], the superiority of KLD over Mutual Information (MI) measure was demonstrated for

2D/3D registration. Learning-based method was further extended to 2D non-rigid image registration in [7] where the KLD w.r.t. a prior joint distribution was minimized together with the maximization of the MI measure.

In this paper, we propose a Jensen-Shannon divergence (JSD) based method for 2D/3D registration. Unlike KLD, JSD is symmetric, bounded, and a true metric, which triggered its popularity in various applications in recent years, ranging from statistical language analysis [8], image edge detection and segmentation [9], to DNA sequence analysis [10]. We notice that JSD provides a more suitable measure than KLD in quantifying histogram discrepancy because some histogram bins may vanish for the training data but not for the observed data or vise versa, in which case KLD is undefined. Furthermore JSD is upper-bounded and symmetric [9], facilitating its easy use as one of the factors in the driving force for registration. Other advantages of our method include: 1) Depending on how well the *a priori* represents the observed data, the registration process is driven by a compounding effect of the statistical consistency of the observed joint histogram to the learned prior and the statistical dependence between the individual intensity distributions of the images being registered; 2) There is no requirement on image segmentation and labeling as done in [6], whose error can lead to further errors in subsequent registration. Instead, an automatic nonlinear histogram mapping is done iteratively during the matching process to handle the intensity discrepancy between the observed data and the training data; 3) An intensity-based histogram is used, which is supposed to result in a higher registration accuracy than [6] where a class-based histogram was used in the argument of a higher computational efficiency in the generation of digitally reconstructed radiographs (DRRs). DRRs are the simulated projection images from the 3D volume that are iteratively produced and compared with the 2D X-ray images during 2D/3D registration. Since DRRs need to be generated a large number of times, the computational efficiency in their calculation directly affects the efficiency of the whole registration algorithm. As will be explained in detail later our fast hardware-based DRR generation makes the utilization of intensity-based histogram totally feasible.

2   Methods

Assume we are given a pair of correctly-registered training images $L = \{l_r, l_f\}$ with a join intensity distribution $p_l(i_r, i_f)$. In the framework of Bayesian inference, the problem of registering a novel pair of images $O = \{o_r, o_f\}$ can be formulated as retrieving the transformation T that maximizes the conditional probability of observing the image pair $O^T = \{o_r, o_f^T\}$ (o_f^T is the floating image o_f after transformation T) given the training pair L and the initially observed pair O:

$$p(T, O^T \mid O, L) \propto p(L \mid T, O^T, O) p(T \mid O^T, O) p(O^T \mid O) \qquad (1)$$

The three terms on the right-hand side of Eq.(1) represent the three factors that drive the matching process. The first term indicates the consistency between the observed pair O^T and the training pair L. The second term specifies the *apriori*-probability of

the transformation T. In the case of rigid-body transformation where all combinations of translations and rotations are considered equally likely, this term can be dropped in maximization. The third term measures the similarity between the two observed images o_r and o_f^T. Hence our method maximizes the compounding effect of the first and the third factors, whose modeling will be articulated in following sections. In comparison, the method in [6] relied only on the first factor for registration, implying that any correspondence between o_r and o_f^T was uniformly favored for registration. This assumption in essence ignores the important fact that two images should have high mutual dependence when registered, which, as will be shown in our experiments, can lead to less robust registration, especially when the observed data deviates from the training data.

2.1 Consistency with Learned Distribution

Assume the observed pair $O^T = \{o_r, o_f^T\}$ for a given transformation T gives rise to a joint histogram $p_o^T(i_r, i_f)$, the first factor in Eq.(1) is modeled by the (inverse) JSD between p_o^T and p_l, stating that independent of the transformation T and the initial observation O, the statistically more coherent the observed histogram to the learned histogram in terms of JSD, the more likely the observed images are registered:

$$p(L|T,O^T,O) = p(L|O^T) \propto -JSD(p_l \parallel p_o^T) \tag{2}$$

$$JSD(p_l \parallel p_o^T) = \frac{1}{2}D_{KL}(p_l \parallel p_M) + \frac{1}{2}D_{KL}(p_o^T \parallel p_M), \; p_M = \frac{1}{2}(p_l + p_o^T) \tag{3}$$

$$\text{and } D_{KL}(p_1 \parallel p_2) = \iint_{i_r,i_f} p_1(i_r,i_f)\log[p_1(i_r,i_f)/p_2(i_r,i_f)]di_r di_f \tag{4}$$

is the well-known KLD measure. It can be shown that unlike KLD, $JSD(p_l \parallel p_o^T)$ in Eq.(3) is upper-bounded by $\log(2)$ (see [9] for proof), and therefore can be easily normalized to be on the same order of magnitude of the third term in Eq.(1). Furthermore, JSD is the square of a true metric that is equivalent to Hellinger metric [9], therefore it is symmetric and removes the nuisance in the use of KLD arising from its asymmetry. We will show that the ordering of the arguments in KLD can yield substantially different values. More importantly, JSD is well-defined even when there exists i_r, i_f such that $p_o^T(i_r, i_f) > 0$ but $p_l(i_r, i_f) = 0$, in which case KLD is undefined. The superiority of JSD over KLD on handling the non-continuity in the learned distribution was demonstrated in [8] via a set of experiments aiming at statistical language analysis.

2.2 Dependence Between the Two Images Being Registered

To model the third factor in Eq.(1), we choose to implement, among numerous similarity measures proposed in the literature, two similarity measures. One is the well-known Mutual Information (MI) and the other is Pattern Intensity (PI), which was reported in [4] to perform the best among six similarity measures tested. Therefore:

$$p(O^T|O) = p(O^T) \propto MI(o_r, o_f^T) \text{ or } \propto PI(o_r, o_f^T) \tag{5}$$

$$MI(o_r, o_f^T) = \iint_{i_r, i_f} p_o^T(i_r, i_f) \log[p_o^T(i_r, i_f) / (p_{o_r}(i_r) p_{o_f}^T(i_f))] di_r di_f \qquad (6)$$

$p_{o_r}(i_r)$ and $p_{o_f}^T(i_f)$ are the marginal distributions for o_r and o_f^T respectively, and

$$PI(o_r, o_f^T) = \iint_{x,y} \sigma^2 / [\sigma^2 + (o_d(x,y) - o_d(v,w))^2] dx dy \qquad (7)$$

Where $o_d = o_r - o_f^T$ is the difference image, $o_d(x,y)$ denotes the intensity value of o_d at pixel (x,y), and $(v-x)^2 + (w-y)^2 < R^2$ defining the region centered at (x,y) with radius R. When two images are registered, there will be a minimum number of structures in the difference image o_d. Pattern intensity considers a pixel to belong to a structure if it has a significantly different intensity value from its neighboring pixels. Good working parameters were shown to be $\sigma = 10$ and $R = 3$ in [4].

2.3 Nonlinear Marginal Histogram Mapping

The marginal intensity distribution of the observed data can differ from that of the training data for many reasons. For example, the window-level for X-ray images are often adjusted during the operation to provide optimal visualization. In digitally subtracted angiography (DSA), the injected contrast agent flows during the sequence, resulting in the difference in the intensity contrast in the subtracted images. In order to maximize the effective range of the learned prior for registration purpose, we suggest a monotonic nonlinear mapping technique to align the marginal intensity distribution of the observed image to that of the training image.

For simplicity, we explain how the reference images o_r and l_r are mapped to each other. The floating images are mapped in the same manner. In the ideal case, intensity i for o_r should be mapped to intensity i' where $C_{o_r}(i) = C_{l_r}(i')$. Here $C_{o_r}(\cdot)$ and $C_{l_r}(\cdot)$ denote the cumulative density function (CDF) of o_r and l_r respectively. For discrete histogram distribution perfect mapping typically can not be achieved and the following algorithm is proposed to achieve an optimal approximation:

$$i \to i' = k : \min_{t_1 \le k \le t_2} \left| p_{o_r}(i) - p_{l_r}(k) \right| \text{ with } t_1 \text{ and } t_2 \text{ defined by:}$$
$$C_{l_r}(t_1 - 1) < C_{o_r}(i-1) \le C_{l_r}(t_1) \text{ and } C_{l_r}(t_2 - 1) < C_{o_r}(i) \le C_{l_r}(t_2) \qquad (8)$$

Here $p_{o_r}(\cdot)$ and $p_{l_r}(\cdot)$ denote the marginal histogram of o_r and l_r respectively. The proposed mapping is similar to the well-known image processing technique called histogram equalization (HE) with two major differences: 1) unlike HE whose targeted distribution is the uniform distribution, our targeted distribution is the marginal histogram of the training image which can be arbitrary; 2) In HE, the mapping is essentially $i \to i' = k$ where $C(k-1) < C_{o_r}(i) \le C(k)$ and $C(\cdot)$ denotes the CDF of the uniform distribution. In contrast, there is a local histogram difference minimization step in our proposed mapping to optimally align the histogram of the observed image to that of the training image, which is crucial for the robust measurement of the similarity between the expected and mapped joint histograms in registration.

2.4   Implementation

The transformation relating points in the 3D volume to points on the projected X-ray image consists of six extrinsic rigid-body parameters $T = \{t_x, t_y, t_z, \vartheta_x, \vartheta_y, \vartheta_z\}$ that are estimated by the iterative registration algorithm, and four intrinsic perspective projection parameters that are determined by the X-ray imaging system based on pinhole cameral model as depicted in Fig.1. DRRs are generated using the 2D texture-based volume rendering technique on the graphics processing unit (GPU), which yields better computational efficiency than software-based DRR generation technique such as ray-casting. It takes about 20ms to generate 256×256 DRRs from a $256 \times 256 \times 253$ volume with an NVidia Quadro FX Go1400, resulting in a typical registration time in the range of 10~30s. At each iteration the X-ray image and the generated DRRs are taken as the observed pair $\{o_r, o_f^T\}$ and are mapped to the training pair $\{l_r, l_f\}$ through nonlinear histogram mapping articulated in Section 2.3. In order to estimate the correct pose T of the 3D volume, Hill-climbing optimization method is implemented to maximize the following variational formulation:

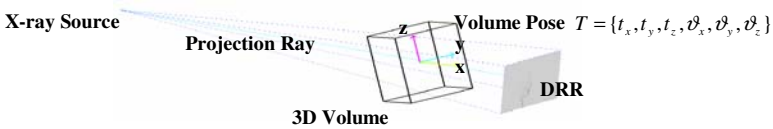$$SM_{total} = \alpha(1 - JSD(p_l \parallel p_o^T)) + (1-\alpha)[MI(o_r, o_f^T) \text{ or } PI(o_r, o_f^T)] \tag{9}$$

Fig. 1. Diagram for the generation of digitally reconstructed radiographs (DRRs)

3   Experiments

The first experiment compares the performance of JSD and KLD on simulated data. In this simplified experiment the reference and floating images in the training pair (Fig.2.a) were exactly the same. The reference image in the testing pair was slightly different from the training reference image in angulations, and the floating image was the horizontally-translated version of the reference image (Fig.2.b). Fig.2.c plots the several measures versus the displacement. We notice that JSD produced a smoother curve than both of the KLD measures, which differed significantly in values due to the switch of the order of the training and testing pairs. The (one minus) MI measure is also plotted as a reference.

   The second experiment investigates the effectiveness of the proposed monotonic nonlinear histogram mapping technique. The training pair was the same as that used in the first experiment, while the testing pair was window-leveled to an apparently different intensity range (Fig.3.a). Fig.3.b plots the marginal histograms of the testing pair, training pair and mapped pair. As shown in Fig.3.c, without mapping, the learned joint histogram could no longer effectively drive registration and the JSD curve became almost uniform across displacements. In contrast, a similar curve as that in the first experiment was achieved after the monotonic nonlinear mapping.
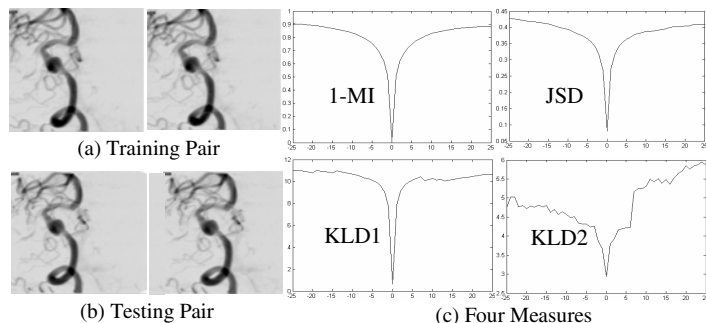
(a) Training Pair

1-MI   JSD

(b) Testing Pair

KLD1   KLD2

(c) Four Measures

Fig. 2. Comparison of four measures versus horizontal displacement: mutual information, JSD, and two KLDs with different orders of the training data and testing data
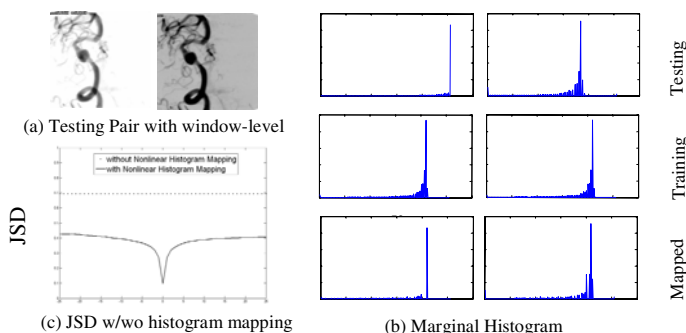

(a) Testing Pair with window-level

(c) JSD w/wo histogram mapping

(b) Marginal Histogram

Testing

Training

Mapped

Fig. 3. Nonlinear histogram mapping to align the marginal histograms of the testing pair with that of the training pair

We further applied our technique to registration of 3D DynaCT to 2D fluoroscopy acquired by a Siemens Axiom Artis System. The object was a chest phantom and the 3D data ($256 \times 256 \times 223$, $0.8 \times 0.8 \times 0.8$mm) was perfectly aligned with the 2D fluoroscopies from two angles. The 2D image from the first angle and the corresponding DRR were used as the training pair (Fig.4.a), and the artificially-moved 3D DynaCT was then registered to the 2D fluoroscopy from the second angle (Fig.4.b). Different types of driving force were tested: prior only (JSD); MI or PI only; and MI + JSD or PI + JSD with $\alpha = 0.5$ in Eq.(9). The whole registration process took about 20s for MI measure and 40s for PI measure on a Pentium®M 2.13G Hz computer. Using prior did not noticeably increase the computational time, partially due to the faster convergence and hence fewer number of iterations. Using PI alone or PI + prior achieved relatively accurate and comparable registration for this data. However, using MI alone resulted in a largely wrong result (Fig.4.). Prior only also did not lead to acceptable registration. In contrast, when combining the prior information with the dependence between the images being registered, much more accurate registration was achieved, suggesting that these two factors provided complimentary information to each other that boosted the robustness of the algorithm.
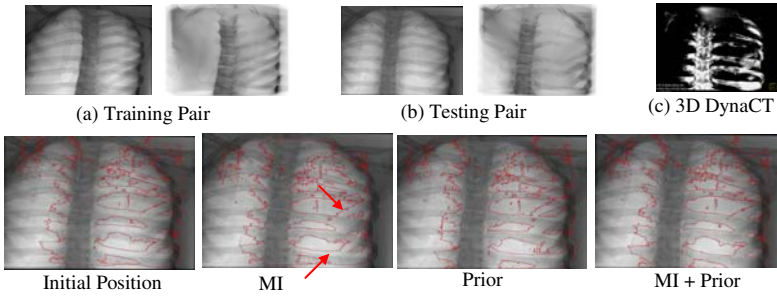
Fig. 4. Registration results of chest phantom using different types of driving force. The outline of the rendered VRT image of the registered 3D volume is overlaid onto the 2D X-ray. Regions pointed by arrows show prominent differences in the accuracy of alignment.
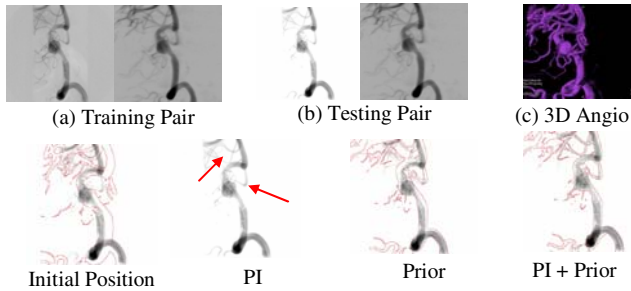

Fig. 5. Registration results of *in vivo* neuro-vascular using different types of driving force

A similar experiment was finally applied to *in vivo* neuro-vascular data acquired during an aneurism operation by a Siemens Axiom Artis dBC biplane system. One frame of the digitally subtracted angiography (DSA) sequence highlighting the contrast agent-filled vessels and the DRR generated from the 3D angio ($256\times256\times253$, $0.3\times0.3\times0.3$mm) after manual alignment were used as the training pair, and the 3D angio without manual alignment was registered to a second frame of the DSA sequence with substantially different contrast intensity (Fig.5.b). It is interesting to notice that there was a wrong global maximum for PI measure, possibly due to the simplicity of the vessel structures present in the image, so that when using PI only the registration was so off as to the outline of the volume did not appear on the 2D image in the result. However, the added factor of the prior information eliminated this wrong global maximum and achieved highly accurate registration when used together with PI. Again prior alone did not produce as accurate registration because of the apparent intensity difference between the two DSA frames used for training and testing. In this experiment MI measure and MI + prior produced comparably accurate results that are not shown. In addition, the fact that the intensity range of the testing DSA frame was substantially different from that of the training frame demonstrated the effectiveness of our histogram mapping technique.

4   Discussion and Conclusion

It has been demonstrated that registration driven by the combined factors of the prior information and the dependence between the images being registered was more robust than when either one of the two factors was used alone. In essence, the learned prior enhanced the registration performance by ruling out the highly unlikely registration that were very different from the training images, and achieved equally good registration in the cases that the conventional similarity measures alone were able to drive a successful registration. The property of being upper-bounded for JSD made the combination of the two factors readily controllable. The nonlinear histogram mapping technique greatly expanded the effective range of the learned prior. More comprehensive experiments are needed to quantitatively evaluate the robustness and accuracy of the proposed method on data from different organs and modalities.

Acknowledgements

References

1. Liu A, Bullitt E, Pizer SM. 3D/2D Registration via Skeletal near Projective Invariance in Tubular Objects. MICCAI'98;952-963.
2. Kita Y, Wilson DL, Nobel JA. Real-Time Registration of 3D Cerebral Vessels to X-Ray Angiograms, MICCAI'98;1125-1133.
3. Hipwell JH, Penny GP, McLaughlin RA, Rhode K, Summers P, Cox TC, Byrne JV, Noble JA, Hawkes DJ. Intensity-Based 2D3D Registration of Cerebral Angiograms. IEEE Trans. on Medical Imaging 2003;22(11):1417-1426.
4. Penny GP, Weese J, Little JA, Desmedt P, Hill DLG, Hawkes DJ. A Comparison of Similarity Measures for Use in 2D3D Medical Image Registration. IEEE Trans. on Medical Imaging 1998;17(4):586-595.
5. Leventon M, Grimson E. Multi-modal Volume Registration Using Joint Intensity Distributions. MICCAI'98;1057-1066.
6. Chan HM, Chung ACS, Yu SCH, Norbash A, Wells WM. Multi-Modal Image Registration by Minimizing Kullback-Leibler Distance between Expected and Observed Joint Class Histograms. CVPR'03;181-190.
7. Guetter C, Xu C, Sauer F, Hornegger J. Learning Based Non-Rigid Multi-Model Image Registration Using Kullback-Leibler Divergence. MICCAI'05;255-262.
8. Lee L. On the Effectiveness of the Skew Divergence for Statistical Language Analysis. Artificial Intelligence and Statistics 2001;65-72.
9. Gomez-Lopera JF. An Analysis of Edge Detection by Using the Jensen-Shannon Divergence. Journal of Mathematical Imaging and Vision 2000;13:35-56.
10. Grosse I et al. Analysis of Symbolic Sequence Using the Jensen-Shannon Divergence. Physical Review E. 2002;65:041905.