Closed-form Jensen-Renyi Divergence for Mixture of Gaussians & Applications to Group-wise Shape Registration *

Fei Wang[1], Tanveer Syeda-Mahmood[1], Baba C. Vemuri[2],
David Beymer[2], and Anand Rangarajan[2]

[1] IBM Almaden Research Center, San Jose, CA, USA
[2] Department of CISE, University of Florida, Gainesville, FL, USA

Abstract. In this paper, we propose a generalized group-wise non-rigid registration strategy for multiple unlabeled point-sets of unequal cardinality, with no bias toward any of the given point-sets. To quantify the divergence between the probability distributions – specifically Mixture of Gaussians – estimated from the given point sets, we use a recently developed information-theoretic measure called Jensen-Renyi (JR) divergence. We evaluate a closed-form JR divergence between multiple probabilistic representations for the general case where the mixture models differ in variance and the number of components. We derive the analytic gradient of the divergence measure with respect to the non-rigid registration parameters, and apply it to numerical optimization of the group-wise registration, leading to a computationally efficient and accurate algorithm. We validate our approach on synthetic data, and evaluate it on 3D cardiac and hippocampal shapes.

1 Introduction

The non-rigid registration of multiple point sets is a fundamental problem in medical imaging for atlas generation and disease class representation. In model-based image segmentation for example, constructing an atlas typically requires us to bring pre-segmented shapes into alignment. Similarly, registration of hippocampal regions across patients with a given disease helps capture the essential disease-specific variations of the region. This is particularly relevant for echocardiography where a disease-specific understanding of echocardiagraphic sequences requires accurate characterization of spatio-temporal motion patterns in the echo videos. The supplemental videos show different patients diagnosed with the same disease of hypokinesia. The similar motion patterns in the designated cardiac regions can be observed. Simultaneous registration of such videos helps summarize the motion pattern for a disease class.

Simultaneous registration of such shapes can be a daunting task due to several reasons. First, the shapes are deformable requiring non-rigid spatial transformations. Secondly, explicit correspondence through search is computationally

---

prohibitive for such large data sets. Robustness to missing and spurious features would be desired. Finally, the registration must be *unbiased* towards any one of the shapes.

In this paper, we develop a novel non-rigid registration technique for multiple point-sets based on a recently introduced information theoretic matching criterion [1] called Jensen-Renyi (JR) divergence. We show that using Renyi's Quadratic entropy results in a closed-form solution for the registration problem. We also derive the analytic gradient of this match measure in order to achieve efficient and accurate non-rigid registration. The Jensen-Renyi measure is then minimized over a class of smooth non-rigid transformations expressed in a thin-plate spline basis. We show the results of our algorithm for the registration of different medical imaging datasets varying from hippocampal region registration to spatio-temporal cardiac echo video registration.

2   Previous Work

The work reported here fits into the general class of approaches that avoid explicit point correspondences for non-rigid registration through the use of divergence measures between probability distributions formed around point sets[2–4]. Specifically, the Jensen-Shannon (JS) divergence and the CDF-based Jensen-Shannon divergence between the feature point-sets was minimized to obtain the non-rigid deformation. The density-based approaches are relatively more robust to the point-sets of different sizes and to the presence of missing features. Furthermore, if an unbiased information theoretic measure is chosen to quantify the multiple densities representing the shapes, the matching results can potentially be unbiased to any of the given point-sets [4].

Other related work includes Wang et al. [5], where the relative entropy measure (Kullback-Leibler divergence) is used to find a similarity transformation between two point-sets under rigid transformation. In Jian et al. [2] and Roy et al. [6], nonrigid registration is between pairs of data sets using L2 distance on a mixture of Gaussians model of the data sets. Both methods, however, have not been extended to the problem of unbiased simultaneous matching of multiple point-sets being addressed in this paper. Recently, in [7], Glaunes et al. represent points as delta functions and match them using the dual norm in a reproducing kernel Hilbert space. The resulting 3D spatial integral is difficult to compute numerically. Similarly, the JS divergence proposed earlier[3] was estimated using the law of large numbers, making it computationally expensive.

In [1], a new divergence measure called the Jensen-Renyi divergence was introduced as a generalization of the Jensen-Shannon divergence being based on Renyi Entropy [1]. In contrast to the JS divergence used earlier [3], the exponential order of Renyi entropy as well as the weights to control the measurement sensitivity of the probability densities can be adjusted in JR divergence. Further, the JR divergence measure enjoys appealing mathematical properties such as convexity and symmetry, affording a great flexibility in a number of applications [1]. Finally, a related work by Chiang et al.[8] has used JR divergence for

pairwise image matching where contiguity constraints between pixels could be utilized.

Our approach differs from prior work in many respects. First, we consider a new problem of simultaneous registration of multiple points sets (rather than multiple pair-wise registrations). Secondly, we introduce the JR divergence for the point-sets registration problem and give a closed form solution to divergence estimation making it computationally attractive. More importantly, Jensen-Renyi divergence can be generalized to a family of divergence measures among a group of point-sets, with JS-divergence being a special case (when $\alpha \to 1$). Finally, our method can accommodate point-sets to be registered of varying size, and is also robust in the presence of noise.

3 Multiple Point-sets Registration using JR Divergence

3.1 Shape Representation by Mixture of Gaussians

Consider first the representation of probability density of shapes as Mixture of Gaussian (MoG). Let the N point-sets to be registered be denoted by $\{X^{(i)}, i \in \{1, ..., N\}\}$. Each point-set $X^{(i)}$ consists of d-dimensional points $\{x_j^{(i)} \in \mathbb{R}^d, j \in \{1, ..., n_p\}\}$. Each point set is represented by a probability density function P_i. The features of ith shape $X^{(i)}$ are represented as a Gaussian Mixture Model

$$
\begin{aligned}
P_i = p(\mathbf{x}|\theta^{(i)}) &= \sum_{a=1}^{K_i} \omega_a^{(i)} \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_a^{(i)}|^{\frac{1}{2}}} \exp\Big(-\frac{1}{2} \big(\mathbf{x} - u_a^{(i)}\big)^T (\Sigma_a^{(i)})^{-1} \big(\mathbf{x} - u_a^{(i)}\big) \Big) \\
&= \sum_{a=1}^{K_i} \omega_a^{(i)} G(\mathbf{x} - u_a^{(i)}, \Sigma_a^{(i)})
\end{aligned}
\tag{1}
$$

where $\mathbf{x} \in \mathbb{R}^d$, parameter set $\theta^{(i)} = \{\omega_a^{(i)}, u_a^{(i)}, \Sigma_a^{(i)}\}$, and $G(\mathbf{x} - u_a^{(i)}, \Sigma_a^{(i)})$ is the Gaussian Kernel in d-dimensional space. Constraints on the MoG weight $\omega_a^{(i)}$ are $\{\omega_a^{(i)} > 0, \sum_{a=1}^{K_i} \omega_a^{(i)} = 1\}$. Since $X_j^{(i)}$ are assumed to be i.i.d., the likelihood of the set of features of $X^{(i)}$ is

$$
p(X^{(i)}|\theta^{(i)}) = \prod_{j=1}^{N_i} \sum_{a=1}^{K_i} \omega_a^{(i)} G(X_j^{(i)} - u_a^{(i)}, \Sigma_a^{(i)})
\tag{2}
$$

For all the shapes, we fit the model parameters $\theta^{(i)}$s by minimizing the negative log-likelihood objective function in Eqn.(2) w.r.t. the model parameters. For simplicity, we specialize to the case where the occupancy probabilities are uniform $\omega_a^{(i)} = \frac{1}{K_i}$ and we have isotropic covariance matrix $\Sigma_a^{(i)} = \sigma_i^2 \mathbf{I}_d$ for all the shapes.

The minimization of Eqn.(2) can be easily accomplished using the well-known EM algorithm for the MoG [9]. This computation is done offline for all the shapes once we fix the number of centroids K_i. Model selection for mixture models needs to be performed to fix the number of centroids.

3.2 A Closed-Form Divergence Measure Between Multiple MoGs

We now derive the cohesive measure between the multiple probabilistic shape representations with the help of the Jensen-Renyi divergence. Please note that distance measures like Jensen-Shannon [3] cannot be derived in closed form for the Gaussian mixture model.

The Jensen-Renyi divergence [1] between probability density functions is defined as:

$$JR_\pi(P_1, P_2, ..., P_N) = H_\alpha\left(\sum \pi_i P_i\right) - \sum \pi_i H_\alpha(P_i), \tag{3}$$

where $\pi = \{\pi_1, \pi_2, ..., \pi_n | \pi_i > 0, \sum \pi_i = 1\}$ are the weights of the probability distributions P_i, $H_\alpha(X)$ is Renyi entropy of order α, $\alpha \geq 0$ defined on a random variable X and is given by $H_\alpha(X) = \frac{1}{1-\alpha} \log(\sum_{i=1}^n p_i^\alpha)$, where p_i are the probabilities of $x_1, x_2, ..., x_n$. Note that Renyi entropy is a generalization of Shannon entropy, since $\lim_{\alpha->1} H_\alpha = H_S$, and H_S is Shannon entropy of the same random variable. Thus Jensen-Shannon divergence is a special case of Jensen-Renyi divergence when $\alpha \to 1$. In the definition of the α-Renyi Entropy, when $\alpha = 2$, $H_2 = -\log(\sum_{i=1}^n p_i^2)$ is called quadratic entropy. We now show that in the case of quadratic entropy ($\alpha = 2$), there exists a closed form solution to the estimation of JR divergence between multiple MoGs, i.e. $JR(p(\mathbf{x}|\theta^{(1)}), ..., p(\mathbf{x}|\theta^{(N)}))$.

Given a Gaussian mixture model $\hat{p}_V(x) = \sum_{i=1}^N \omega_i G(x - v_i, \sigma^2\mathbf{I})$, where v_i and $\sigma^2\mathbf{I}$ are the centroids and covariance matrix respectively, unlike Shannon entropy, Renyi's quadratic entropy has a closed form for MoGs

$$H_2(V) = -\log \int_{-\infty}^{+\infty} \hat{p}_V(x)^2 dx = -\log\Big[\sum_{j=1}^N \int_{-\infty}^{+\infty} \omega_i\omega_j G(x - v_i, \sigma^2\mathbf{I})$$
$$G(x - v_j, \sigma^2\mathbf{I})dx\Big] = -\log\Big[\sum_{i=1}^N \sum_{j=1}^N \omega_i\omega_j G(v_i - v_j, 2\sigma^2\mathbf{I})\Big], \tag{4}$$

where $v_i \in \mathbb{R}^d$ and $v_j \in \mathbb{R}^d$ are two centroids, and Σ_i and Σ_j are two covariance matrices for two Gaussian kernels in the space. Substituting the value of quadratic entropy in Eqn.(3) for the MoG of each shape expressed in Eqn.(1), we get an estimate for the quadratic entropy of each individual probabilistic shape representation

$$H_2(P_i) = -\log\Big[\frac{1}{K_i^2} \sum_{k=1}^{K_i} \sum_{l=1}^{K_i} G(u_k^{(i)} - u_l^{(i)}, 2\sigma_i^2\mathbf{I})\Big]. \tag{5}$$

For the convex combination $\sum \pi_i P_i$, if we choose $\pi_i = \frac{K_i}{M}$, where $M = \sum_i K_i$ is the total number of the centroids in all shapes, we have the following,

$$\sum_{i=1}^N \pi_i P_i = \frac{1}{M} \sum_{i=1}^N \sum_{a=1}^{K_i} G(\mathbf{x} - u_a^{(i)}, \sigma_i^2\mathbf{I}) = \frac{1}{M} \sum_{j=1}^M G(\mathbf{x} - u_j, \sigma_{\tau(j)}^2\mathbf{I}) \tag{6}$$

where $\{u_1, u_2, ..., u_M\} \equiv \{u_1^1, ..., u_j^i, ..., u_{K_N}^N\}$ are the pooled centroids, $\tau : \{1, ..., M\} \to \{1, ..., N\}$ is a mapping function that maps the index of an individual point to the index of the point-set. Therefore the linear combination of the MoGs can be expressed as a single Gaussian Mixture centered on the

pooled centroids. Consequently, we have the quadratic entropy estimation of the $\sum \pi_i P_i$,

$$H_2(\sum_{i=1}^{N} \pi_i P_i) = -\log\Big[\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} G(u_i - u_j, (\sigma_{\tau(i)}^2 + \sigma_{\tau(j)}^2)\mathbf{I})\Big] \tag{7}$$

Combining the two terms in Eqn.(5) and (7) together, we have the closed-form JR-divergence for multiple MoGs,

$$JR_\pi(P_1, P_2, ..., P_N) = H_2(\sum \pi_i P_i) - \sum \pi_i H_2(P_i) = -\log\Big[\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}$$

$$G(u_i - u_j, (\sigma_{\tau(i)}^2 + \sigma_{\tau(j)}^2)\mathbf{I})\Big] + \sum_{i=1}^{N}\frac{K_i}{M}\log\Big[\frac{1}{K_i^2}\sum_{k=1}^{K_i}\sum_{l=1}^{K_i} G(u_k^i - u_l^i, 2\sigma_i^2\mathbf{I})\Big] \tag{8}$$

The JR divergence can be further simplified using matrix notation. Let us introduce a Gaussian kernel matrix $\mathbf{G}$ with $G_{ij} = G(u_i - u_j, (\sigma_{\tau(i)}^2 + \sigma_{\tau(j)}^2)\mathbf{I})$, and define an indicator vectors I_a (of length M) for a^{th} point-set, i.e. $I_a(i) = 1$ if u_i is from the ath point-set, i.e. $\tau(i) = a$. I_M is the vector whose elements are all ones. Eqn. (8) can be rewritten as

$$JR_\pi(P_1, P_2, ..., P_N) = -\log\frac{I_M^T \mathbf{G} I_M}{M^2} + \sum_i \frac{K_i}{M}\log\frac{I_i^T \mathbf{G} I_i}{K_i^2} \tag{9}$$

This is the final expression for the distance function used in this paper. The number of centroids and the variances of each shape can be different.

3.3 Cost Function Optimization

We now turn to the description of the deformation model. We assume that parameters of each shape $\theta^{(i)} = \{u^{(i)}, \sigma_i\}$ (comprising the centroids and variance) are simultaneously deforming until the JR divergence between $\{p(\mathbf{x}|\theta^{(i)}), i \in \{1, \ldots, N\}\}$ are minimized. We use the familiar thin-plate spline (TPS) deformation model for the centroids, for which we refer the readers to [3] for more details. In addition a regularization term is added to JR divergence to prevent reflections and unphysical affine transformations.

Based on Eqn.(9), we can derive the gradient of the JR divergence with respect to the transformation parameters $\boldsymbol{\mu}^{(a)}$, which is given by,

$$\frac{\partial JR}{\partial \boldsymbol{\mu}^a} = -\frac{I_M^T \frac{\partial \mathbf{G}}{\partial \boldsymbol{\mu}^a} I_M}{I_M^T \mathbf{G} I_M} + \sum_{i=1}^{N}\frac{K_i}{M}\frac{I_i^T \frac{\partial \mathbf{G}}{\partial \boldsymbol{\mu}^a} I_i}{I_i^T \mathbf{G} I_i} \tag{10}$$

The details of the derivation are omitted due to space limitations. Once we have the analytical gradient, the cost function optimization is achieved very efficiently using a quasi-Newton method. The variance parameters σ_i are updated separately from the remaining parameters and are not allowed to abruptly change.

4 Results

We now present experimental results on both synthetic and real data sets. We first demonstrate the robustness and accuracy that can be achieved in registration with JR divergence in comparison to Jensen-Shannon divergence reported in an earlier work. We then report results of registration on two data sets drawing from cardiac imaging and brain imaging.

4.1 Robustness to Noise and Outliers

We generated a random 2D point set for our first set of experiments. To test the presence of noise and outliers, we corrupted the model set of size n (1) by discarding a subset of size $(1 - \rho)n$ from the model point set, (2) by applying a rigid transformation $(\mathbf{R}, \mathbf{t})$ to the template, (3) by perturbing the points of the template with noise (of strength ϵ), and (4) by adding $(\tau - \rho)n$ spurious, uniformly distributed points to the template. Thus, after corruption, a template point set will have a total of τn points, of which only ρn correspond to points in the model set. At each of several noise levels and outliers strengths, we generate five models and six corrupted templates from each model for a total of 30 pairs at each noise and outlier strength setting. For each pair, we use our algorithm and the Jensen-Shannon method to estimate the known rigid transformation which was partially responsible for the corruption. Results show that our method exhibits stronger resistance to outliers than the JS method, as shown in Table 1. Furthermore, the average running time for all synthetic examples is $4.61s$ in our method compared with $5.71s$ in the JS algorithm (single core, 1.66 GHz processor).

JR improvement over JS in $\mathbf{R}, \mathbf{t}$ estimation (%)	outlier fraction $(\tau - \rho)/(\rho)$				
	0.1	0.2	0.3	0.4	0.5
rotation $\mathbf{R}$ improvement (%)	11.23%	13.57%	4.36%	27.91%	14.90%
translation $\mathbf{t}$ improvement (%)	5.26%	7.24%	3.55%	-3.53%	13.83%

Table 1. Robustness to outliers in the presence of large noise.

4.2 Evaluation on 3D Shapes from Cardiac Echo Videos

Next, we presents results over 3D shapes extracted from cardiac echo videos. Specifically, cardiologists from hospitals in India provided us echo videos of patients with various diseases, which depicted over 500 heart beat cycles chosen from over 50 patients with a number of cardiac diseases including myocardial infarction, cardiomyopathy, mitral regurgitation, and regional wall motion abnormality etc. For each disease class, we collected videos depicting similar views (short axis, long axis, four chamber views). An Active Shape Model (ASM) was
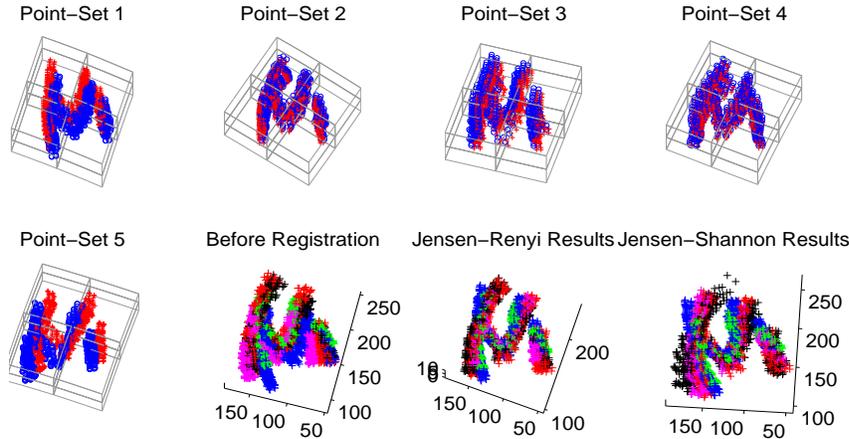
Fig. 1. Experiment results on five 3D echocardiogram point-sets. (see text for details).

used to characterize each such view as described in [10]. Feature points corresponding to identifiable landmarks on heart wall boundaries were automatically extracted and tracked [10] to obtain a 3D point set.

We showcase results on 3D shapes for five patients of Parasternal Long Axis (PLA) views, which are formed by stacking together the points from all frames of a single cycle, and the time axis is normalized with respect to the cycle. As shown in Fig. 1, the recovered deformation between each point-set (blue) and the mean shape (red) are superimposed on the first row and the left image in second row in Figure 1. The point-sets before and after registration results are shown in second and third image of the second row of Figure 1. The registration results generated using Jensen-Shaonnon is shown in the lower-right for comparison, from which we can observe that the results generated using our algorithm exhibits more similarity than the JS approach. As we described earlier, all these results are computed simultaneously and automatically. This example demonstrates that our matching algorithm can simultaneously align multiple shapes even thought the information on point correspondences is not in anyway available.

The advantage of JR over JS in registering point sets also exists within disease categories. In Fig 2, we show the remaining point variance after registration of videos from a number of diseases: regional wall motion abnormality (RWMA), mitral regurgitation (MR), and myocardial infarction (MI). For all diseases, the JR variance is lower than the corresponding JS variance, showing that JR performs a superior registration of the point sets.

5 Conclusions

In this paper, we presented a novel and robust algorithm that utilizes an information theoretic measure, namely Jensen-Renyi divergence, to simultaneously
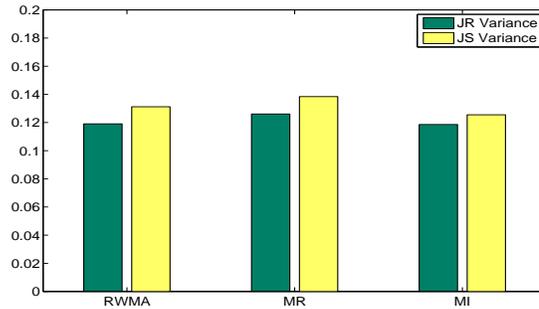
Fig. 2. Within different cardiac diseases, registration variance is lower with our JR approach compared with JS, showing we achieve a better overall registration.

register multiple unlabeled point-sets. Using the quadratic form of Renyi entropy, we have shown that it is possible to obtain a closed form solution to the non-rigid registration problem leading to computational efficiency in registration. While we used Gaussian kernels to represent the probability density of point sets, the formalism holds for other kernels as well. Experiments were depicted with both 2D and 3D point sets from medical and non-medical domains. Future work will focus on generalizing the non-rigid deformations to diffeomorphic mappings.

References

1. Hamza, A.B., Krim, H.: Jensen-renyi divergence measure: theoretical and computational perspectives. ISIT (2003) 257–257
2. Jian, B., Vemuri, B.: A robust algorithm for point set registration using mixture of gaussians. In: ICCV2005. (2005) 1246–1251
3. Wang, F., Vemuri, B.C., Rangarajan, A., Eisenschenk, S.J.: Simultaneous nonrigid registration of multiple point sets and atlas construction. PAMI (2008) 2011–2022
4. Wang, F., Vemuri, B.C., Rangarajan, A.: Groupwise point pattern registration using a novel CDF-based jensen-shannon divergence. In: CVPR. (2006) 1283–1288
5. Wang, Y., Woods, K., McClain, M.: Information-theoretic matching of two point sets. IEEE Transactions on Image Processing **11**(8) (August 2002) 868–872
6. Roy, A., Gopinath, A., Rangarajan, A.: Deformable density matching for 3d non-rigid registration of shapes. In: MICCAI. (2007) 942–949
7. Glaunes, J., Trouvé, A., Younes, L.: Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching. In: CVPR. (2004) 712–718
8. Chiang, M.C., Dutton, R.A., Hayashi, K.M., Toga, A.W., Lopez, O.L., Aizenstein, H., Becker, J.T., Thompson, P.M.: Fluid registration of medical images using jensen-renyi divergence reveals 3d profile of brain atrophy in HIV/AIDS. ISBI (2006) 193–196
9. Yuille, A.L., Stolorz, P., Utans, J.: Statistical physics, mixtures of distributions, and the EM algorithm. Neural Comput. **6**(2) (1994) 334–340
10. Syeda-Mahmood, T., Wang, F., Beymer, D., London, M., Reddy, R.: Characterizing spatio-temporal patterns for disease discrimination in cardiac echo videos. In: MICCAI. (2007) 261–269