A generalized divergence measure for robust image registration

Article *in* IEEE Transactions on Signal Processing · June 2003

DOI: 10.1109/TSP.2003.810305 · Source: IEEE Xplore

3 authors, including:

Hamid Krim
North Carolina State University
261 PUBLICATIONS **4,849** CITATIONS

A Generalized Divergence Measure for Robust Image Registration

Yun He, A. Ben Hamza, and Hamid Krim, *Senior Member, IEEE*

***Abstract*—Entropy-based divergence measures have shown promising results in many areas of engineering and image processing. In this paper, we define a new generalized divergence measure, namely, the *Jensen–Rényi* divergence. Some properties such as convexity and its upper bound are derived. Based on the Jensen–Rényi divergence, we propose a new approach to the problem of image registration. Some appealing advantages of registration by Jensen–Rényi divergence are illustrated, and its connections to mutual information-based registration techniques are analyzed. As the key focus of this paper, we apply Jensen–Rényi divergence for inverse synthetic aperture radar (ISAR) image registration. The goal is to estimate the target motion during the imaging time. Our approach applies Jensen–Rényi divergence to measure the statistical dependence between consecutive ISAR image frames, which would be maximal if the images are geometrically aligned. Simulation results demonstrate that the proposed method is efficient and effective.**

***Index Terms*—Image registration, information divergence, inverse SAR imaging, Rényi entropy.**

I. Introduction

IMAGE registration is an important problem in computer vision [1], [2], remote sensing data processing [3], [4], and medical image analysis [5], [6]. The key step of image registration is to find a spatial transformation such that a similarity metric between two or more images taken at different times, from different sensors or from different viewpoints, achieves its maximum.

One such example, which is of primary interest in the sequel, is inverse synthetic aperture radar (ISAR) imaging. ISAR is a microwave imaging system capable of producing high-resolution imagery from data collected by a relatively small antenna. The ISAR imaging is induced by target motion, which unfortunately also blurs the resulting image. After a standard translational focusing process, image registration can be applied to estimate the target rotational motion parameter, on which polar re-formatting may be used to achieve yet a higher resolution image. Related work in this area includes image registration in interferometric SAR processing by Gabriel [7], Li [8], and Lin [9] and Fornaro [10].

Manuscript received April 5, 2001; revised November 26, 2002. This work was supported by the AFOSR under Grant F49620-98-1-0190 and by the ONR-MURI under Grant JHU-72798-S2. The associate editor coordinating the review of this paper and approving it for publication was Prof. Randolph L. Moses.

Y. He is with Tality Corporation, Cary, NC 27511 USA (e-mail: yhe@tality.com).

A. Ben Hamza and H. Krim are with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695 USA.

Digital Object Identifier 10.1109/TSP.2003.810305

Over the last three decades, a wide variety of registration techniques have been developed for different applications. These techniques may be classified [11] into correlation methods, Fourier methods, landmark mapping, and elastic model-based matching.

Given two images f_1, f_2: $\Omega \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ (Ω is a bounded set, and it is usually a rectangle), correlation methods [12] calculate the normalized two-dimensional (2-D) cross-correlation function $C(f_1, \mathcal{T}_{(\ell,\theta,\gamma)}f_2)$ between f_1 and f_2, where $\mathcal{T}$ is a Euclidean transformation with translational parameter $\ell = (\ell_x, \ell_y)$, a rotational parameter θ , and a scaling parameter γ. The registration problem may then be succinctly stated as

$$(\ell^*, \theta^*, \gamma^*) = \arg \max_{(\ell,\theta,\gamma)} C\left(f_1, \mathcal{T}_{(\ell,\theta,\gamma)}f_2\right). \tag{1}$$

The correlation methods are generally limited to registration problems in which the image is misaligned by only a small rigid transformation. In addition, the peak of the correlation may not be clearly discernible in the presence of noise. Fourier methods [13] are the frequency domain equivalent of the correlation methods. Fourier methods make use of the translational property of the Fourier transform and search for the optimal spectral match between two images. Since rotation is invariant under a Fourier transformation, rotating an image merely rotates the Fourier transform of that image [14]. If we denote F_1, F_2 as the 2-D Fourier transforms of f_1, f_2, respectively, we obtain the phase of the cross-power spectrum rotated by θ as

$$P_\theta(\omega_x, \omega_y) = \frac{F_1(\omega_x,\omega_y)\mathcal{T}_\theta F_2(\omega_x,\omega_y)}{|F_1(\omega_x,\omega_y)\mathcal{T}_\theta F_2(\omega_x,\omega_y)|}. \tag{2}$$

To determine the rotational parameter θ, one proceeds to maximize the 2-D inverse Fourier transformation of $P_\theta(\omega_x, \omega_y)$, that is, a cross-correlation that is as peaked or as impulsive as possible, and the location of that impulse is exactly the translational parameter ℓ. In light of their equivalence to the correlation methods, Fourier methods are also limited to registration problems with a small rigid transformation. If there exists spatially local variation, then both the correlation methods and the Fourier methods would fail. For cases of unknown misalignment type, landmark mapping techniques [15] and elastic model-based matching [16], [17] may be used to tackle the registration problem. Landmark mapping techniques extract feature points from a reference image and a target image, respectively, and then apply a piecewise interpolation to compute a transformation for mapping the feature point sets from the reference image to the target image. Landmark-based methods are usually computationally intensive, and their

accuracy depends on the degree of reliability of the feature points. Instead of finding the mapping between the feature point sets, elastic model-based matching methods model the distortion in the image as the deformation of an elastic material. The resulting registration transformation is the deformation with a minimal bending and stretching energy. Practical elastic model-based methods [18] are also based on computationally expensive iterative algorithms, and the choice of feature points plays a crucial role in their performance.

In the work of Woods [19] and Viola [20], mutual information, which is a basic concept from information theory, is introduced as a measure for evaluating the similarity between images. When the two images are properly aligned, corresponding areas overlap, and the resulting joint histogram exhibits high values for the pixel combinations of the corresponding regions. When the images are misregistered, nonmatched areas also overlap and will contribute to additional pixel combinations in the joint histogram. In case of misregistration, the joint histogram has fewer significant peaks and is more dispersed than that of the correct alignment of images. The registration criterion is hence to find a transformation such that the mutual information of the corresponding pixel pair intensity values in the matching images is maximized. This approach is widely accepted [21] as one of the most accurate and robust registration techniques. Following the same argument, Hero *et al.* [22] extend this approach by applying Rényi entropy to measure the joint histogram as a similarity metric between images. On the other hand, Fisher *et al.* propose mutual information based approaches to feature extraction for ATR [23] as well as to the analysis of functional MRI data [24].

Inspired by this previous work and looking to address their limitation in often difficult imagery, we introduce in this paper a novel generalized information theoretic measure: a *Jensen–Rényi* divergence that we define in terms of Rényi entropy [25]. Jensen–Rényi divergence is defined as a similarity measure among any finite number of weighted probability distributions. Shannon mutual information is a limiting case of the Jensen–Rényi divergence. This generalization provides us with an ability to control the measurement sensitivity of spatial dependency and, hence, ultimately results in a better registration accuracy.

In the next section, we give a brief description of the problem, which motivated and yielded this investigation. In Section III, we introduce the Jensen–Rényi divergence and its properties. Section IV describes the concepts of image registration with the Jensen–Rényi divergence. Numerical experiments for ISAR image registration is demonstrated in Section V. We finally provide concluding remarks in Section VI.

II. Problem Statement

ISAR imagery represents reflectivity magnitude associated with the illuminated target. The mechanism of ISAR can be explained in terms of spotlight SAR [26], as illustrated in Fig. 1. Spotlight SAR is the result of a radar antenna constantly tracking a particular target of interest. The same data would be collected if the radar were stationary and the target were rotating. This very rotation is used to generate the target image

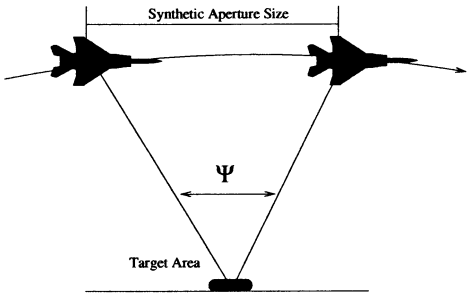

Fig. 1. Spotlight SAR.

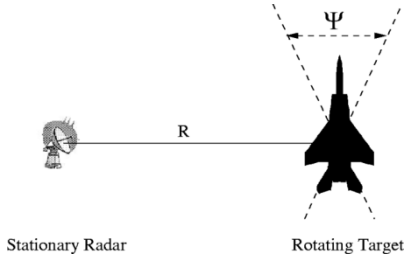

Fig. 2. SAR/ISAR equivalence: ISAR geometry.

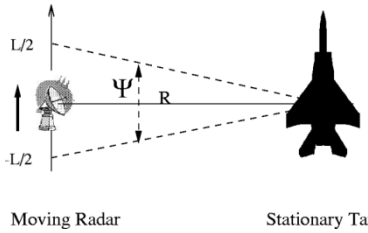

Fig. 3. SAR/ISAR equivalence: SAR geometry.

and constitutes the essence of ISAR. In the terminology of radar signal processing, the direction of radar line of sight (LOS) is referred to as *range*, and the direction orthogonal to range is referred to as *cross-range* or *azimuth*. Fig. 2 illustrates the data collection of an airborne target rotating through an angle Ψ. The spotlight SAR equivalent geometry is the moving radar in Fig. 3, which collects the same data while flying a circular segment around an identical but nonrotating target. The SAR aperture length L in Fig. 3 corresponds to the integration angle Ψ in Fig. 2. The resolution of ISAR imagery is defined in terms of range resolution and cross-range resolution. The range resolution δr is determined by the bandwidth of the transmitted radar pulse β and given by [27]

$$\delta r = \frac{c}{2\beta} \tag{3}$$

where c is the speed of light. The cross-range resolution of ISAR imaging system is given by [27]

$$\delta r_c = \frac{\lambda}{2\Psi} \tag{4}$$

where λ is the carrier wavelength. See [26] for a more thorough discussion of ISAR.

By (4), it is clear that the larger the integration angle, the better the cross-range resolution. However, ISAR imagery is induced by target motion, and the target motion in turn causes

time-varying spectra of the received signals. Motion compensation has to be applied to obtain a high-resolution image. The objective of ISAR image registration is to estimate the target motion during the imaging time. Let $\mathcal{T}_{(\ell,\theta,\gamma)}$ be a Euclidean transformation with a translational parameter $\ell = (\ell_x, \ell_y)$, a rotational parameter θ, and a scaling parameter γ. Given two ISAR image frames f_1 and f_2, the estimates of target motion parameters $(\ell^*, \theta^*, \gamma^*)$ are given by

$$(\ell^*, \theta^*, \gamma^*) = \arg \max_{(\ell,\theta,\gamma)} S_d\left(f_1, \mathcal{T}_{(\ell,\theta,\gamma)} f_2\right) \tag{5}$$

where S_d is a measure induced by a given metric d. This induced measure is maximal if f_1 matches $\mathcal{T}_{(\ell,\theta,\gamma)} f_2$. As the radar tracks a target, the reflected signal is continuously recorded during the imaging time. By registering a sequence of consecutive image frames $\{f_i\}_{i=0}^N$, the target motion during the imaging time can be estimated by interpolating $\{(\ell_i, \theta_i, \gamma_i)\}_{i=1}^N$. Based on the estimated trajectory of the target, translational motion compensation (TMC) and rotational motion compensation (RMC) [26] can be used to generate a focused image of the target.

III. Jensen–Rényi Divergence

Let $k \in \mathbb{N}$ and $\mathcal{X} = \{x_1, x_2, \ldots, x_k\}$ be a finite set with a probability distribution $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)$, i.e., $\sum_{j=1}^k p_j = 1$ and $p_j = P(x_j) \geq 0$, where $P(\cdot)$ denotes the probability.

Rényi entropy is a generalization of Shannon entropy [28] and is defined as

$$R_\alpha(\boldsymbol{p}) = \frac{1}{1-\alpha} \log \sum_{j=1}^k p_j^\alpha, \qquad \alpha > 0 \text{ and } \alpha \neq 1. \tag{6}$$

For $\alpha > 1$, the Rényi entropy is neither concave nor convex.

For $\alpha \in (0, 1)$, it is easy to see that Rényi entropy is concave, and tends to Shannon entropy $H(\boldsymbol{p})$ as $\alpha \to 1$ [25]. It can easily be verified that R_α is a nonincreasing function of α, and hence

$$R_\alpha(\boldsymbol{p}) \geq H(\boldsymbol{p}), \qquad \forall\, \alpha \in (0, 1). \tag{7}$$

In the sequel, we will restrict $\alpha \in (0, 1)$, unless otherwise specified, and will use a base 2 logarithm, i.e., the measurement unit is *bits*.

As shown in Fig. 4, the measure of uncertainty is at a minimum when Shannon entropy is used, and it increases as α decreases. Rényi entropy attains a maximum uncertainty when α is equal to zero.

Definition 1: Let $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$ be n probability distributions on $\mathcal{X}$, and let $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)$ be a weight vector such that $\sum_{i=1}^n \omega_i = 1$ and $\omega_i \geq 0$. We define the Jensen–Rényi divergence as

$$\mathrm{JR}_\alpha^{\boldsymbol{\omega}}(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n) = R_\alpha\left(\sum_{i=1}^n \omega_i \boldsymbol{p}_i\right) - \sum_{i=1}^n \omega_i R_\alpha(\boldsymbol{p}_i)$$

where $R_\alpha(\boldsymbol{p})$ is the Rényi entropy, $\alpha > 0$, and $\alpha \neq 1$.

Using the Jensen inequality, it is easy to check that the Jensen–Rényi divergence is non-negative for $\alpha \in (0, 1)$. It is also symmetric and vanishes if and only if the probability distributions $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$ are equal, for all $\alpha > 0$. Fig. 5 illustrates the three-dimensional (3-D) representation of the Jensen–Rényi divergence for two Bernoulli probability distributions, with $\alpha = 0.5$.

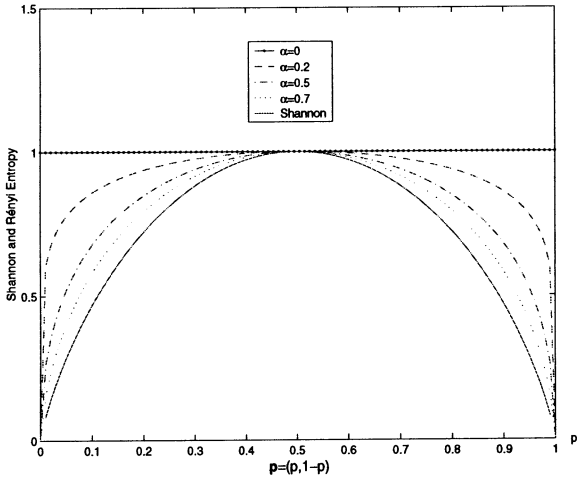

Fig. 4. Shannon and Rényi entropy of Bernoulli distribution $\boldsymbol{p} = (p, 1-p)$ for different values of α.

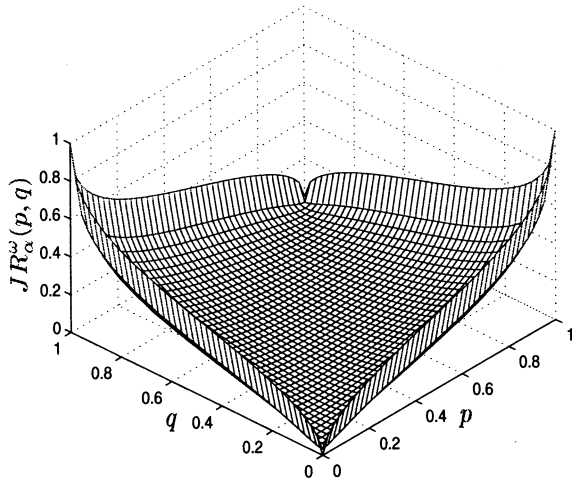

Fig. 5. Three-dimensional representation of Jensen–Rényi divergence $JR_\alpha^{\boldsymbol{\omega}}(\boldsymbol{p}, \boldsymbol{q}), \boldsymbol{p} = (p, 1-p), \boldsymbol{q} = (q, 1-q), \alpha = 0.5, \boldsymbol{\omega} = (0.5, 0.5)$.

When $\alpha \to 1$, the Jensen–Rényi divergence is exactly the generalized Jensen–Shannon divergence [29].

Unlike other entropy-based divergence measures such as the well-known Kullback divergence, the Jensen–Rényi divergence has the advantage of being symmetric and generalizable to any finite number of probability distributions, with a possibility of assigning weights to these distributions.

The following result establishes the convexity of the Jensen–Rényi divergence of a set of probability distributions.

Proposition 1: For $\alpha \in (0, 1)$, the Jensen–Rényi divergence $JR_\alpha^{\boldsymbol{\omega}}$ is a convex function of $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$.

Proof: See the Appendix. ■

The following result, in a sense, clarifies and justifies calling upon the Jensen–Rényi divergence as a measure of disparity among probability distributions.

Proposition 2: The Jensen–Rényi divergence achieves its maximum value when $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$ are degenerate distributions, that is, $\boldsymbol{p}_i = \{\delta_{ij}\}$, where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

Proof: Denote by $\boldsymbol{p}_i = \{p_{ij}\}$ the ith probability distribution, where $1 \le i \le n$ and $1 \le j \le k$. Clearly, the domain of $JR_\alpha^{\boldsymbol{\omega}}$ is $[0, 1]^{nk}$. This domain is a convex polytope [30] in which the vertices are degenerate probability distributions. That is, the maximum value of the Jensen–Rényi divergence occurs at one of the degenerate distributions. ■

Since the Jensen–Rényi divergence is a convex function of $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$, it achieves its maximum value when the Rényi entropy function of the $\boldsymbol{\omega}$-weighted average of degenerate probability distributions achieves its maximum value as well.

Assigning weights ω_i to the degenerate distributions $\Delta_1, \Delta_2, \ldots, \Delta_n$, $\Delta_i = \{\delta_{ij}\}$, $j = 1, 2, \ldots k$, the following upper bound

$$JR_\alpha^{\boldsymbol{\omega}} \le R_\alpha\left(\sum_{i=1}^{n} \omega_i \Delta_i\right) \tag{8}$$

which easily falls out of the Jensen–Rényi divergence, may be used as a starting point. Without loss of generality, consider the Jensen–Rényi divergence with equal weights $\omega_i = 1/n$ for all i, and denote it simply by JR_α, to write

$$\begin{aligned} JR_\alpha &\le R_\alpha\left(\sum_{i=1}^{n} (\Delta_i/n)\right) \\ &= \frac{1}{1-\alpha} \log \sum_{j=1}^{k} \left(\sum_{i=1}^{n} (\delta_{ij}/n)\right)^{\alpha} \\ &= R_\alpha(\boldsymbol{a}) + \frac{\alpha}{\alpha - 1} \log(n) \end{aligned} \tag{9}$$

where

$$\boldsymbol{a} = (a_1, a_2, \ldots, a_k) \quad \text{such that} \quad a_j = \sum_{i=1}^{n} \delta_{ij}. \tag{10}$$

Since $\Delta_1, \Delta_2, \ldots, \Delta_n$ are degenerate distributions, then we have $\sum_{j=1}^{k} a_j = n, \forall k \ge n$. From (9), it is clear that JR_α achieves its maximum value when $R_\alpha(\boldsymbol{a})$ does as well.

In order to maximize $R_\alpha(\boldsymbol{a})$, the concept of majorization [31] will be used. Let $(x_{[1]}, x_{[2]}, \ldots, x_{[k]})$ denote a nonincreasing ordering of the components of a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)$.

Definition 2: Let $\boldsymbol{a}$ and $\boldsymbol{b} \in \mathbb{R}^k$, where $\boldsymbol{a}$ is said to be majorized by $\boldsymbol{b}$, written $\boldsymbol{a} \prec \boldsymbol{b}$, if

$$\begin{cases} \displaystyle\sum_{j=1}^{k} a_{[j]} = \sum_{j=1}^{k} b_{[j]} \\ \displaystyle\sum_{j=1}^{\ell} a_{[j]} \le \sum_{j=1}^{\ell} b_{[j]}, \quad \ell = 1, 2, \ldots, k-1. \end{cases}$$

Definition 3: A real-valued function ϕ defined on a set $\Omega \subset \mathbb{R}^k$ is said to be Schur-concave on Ω if

$$\boldsymbol{a} \prec \boldsymbol{b} \implies \phi(\boldsymbol{a}) \ge \phi(\boldsymbol{b}), \quad \forall \boldsymbol{a}, \boldsymbol{b} \in \Omega.$$

Define the function $g(a_j) = a_j^\alpha$, $\alpha \in (0, 1)$ on an interval $J \subset \mathbb{R}$. It is clear that g is a concave function on J; thus, $\phi(\boldsymbol{a}) = \sum_{j=1}^{k} g(a_j)$ is Schur-concave [31] on J^k, that is

$$\boldsymbol{a} \prec \boldsymbol{b} \implies \sum_{j=1}^{k} g(a_j) \ge \sum_{j=1}^{k} g(b_j).$$

Since $\log(\cdot)$ is an increasing function, and $\alpha \in (0, 1)$, it follows that

$$\boldsymbol{a} \prec \boldsymbol{b} \implies R_\alpha(\boldsymbol{a}) \ge R_\alpha(\boldsymbol{b}).$$

Therefore, $R_\alpha(\cdot)$ is a Schur-concave function. The following result establishes the maximum value of the Jensen–Rényi divergence.

Proposition 3: Let $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$ be n probability distributions with

$$\boldsymbol{p}_i = (p_{i1}, p_{i2}, \ldots, p_{ik}), \quad \sum_{j=1}^{k} p_{ij} = 1, \quad p_{ij} \ge 0.$$

If $k \equiv r \pmod n$, $0 \le r < n$, then

$$\mathrm{JR}_\alpha \le \frac{1}{1-\alpha} \log(k - r) \tag{11}$$

where $\alpha \in (0, 1)$.

Proof: It is clear that the vector

$$\boldsymbol{g} = (\overbrace{n/(k-r), \ldots, n/(k-r)}^{k-r}, \overbrace{0, \ldots, 0}^{r})$$

is majorized by the vector $\boldsymbol{a}$ defined in (10). Hence, $R_\alpha(\boldsymbol{a}) \le R_\alpha(\boldsymbol{g})$. Invoking (9) completes the proof. ■

According to Proposition 3, and for the special case of $k \equiv 0 \pmod n$, the following inequality holds:

$$\mathrm{JR}_\alpha(\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n) \le \log(k).$$

IV. Image Registration With Jensen–Rényi Divergence

Let f_1, f_2 be two digital images defined on a bounded domain $\Omega \subseteq \mathbb{N}^2$, where the goal of image registration in the context of the Jensen–Rényi divergence is to determine the spatial transformation parameters $(\ell^*, \theta^*, \gamma^*)$ such that

$$\begin{aligned} (\ell^*, \theta^*, \gamma^*) &= \arg\max_{(\ell,\theta,\gamma)} S_d\left(f_1, \mathcal{T}_{(\ell,\theta,\gamma)} f_2\right) \\ &= \arg\max_{(\ell,\theta,\gamma)} JR_\alpha^{\boldsymbol{\omega}}(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n) \end{aligned} \tag{12}$$

where $\boldsymbol{p}_i = \boldsymbol{p}_i(f_1, \mathcal{T}_{(\ell,\theta,\gamma)} f_2)$, $1 \le i \le n$, and the measure S_d defined in (5) is induced from a Jensen–Rényi divergence of order α and weight $\boldsymbol{\omega}$.

Denote $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ and $\mathcal{Y} = \{y_1, y_2, \ldots, y_n\}$ as the sets of pixel intensity values of f_1 and $\mathcal{T}_{(\ell,\theta,\gamma)} f_2$, respectively, and let X, Y be two random variables taking values in $\mathcal{X}$ and $\mathcal{Y}$. $\boldsymbol{p}_i(f_1, \mathcal{T}_{(\ell,\theta,\gamma)} f_2) = (p_{ij})_{1 \le j \le n}$ is defined as

$$p_{ij} = P(Y = y_j | X = x_i), \qquad j = 1, 2, \ldots, n$$

which is the conditional probability of $\mathcal{T}_{(\ell,\theta,\gamma)} f_2$ given f_1 for the corresponding pixel pairs. Here, the Jensen–Rényi divergence acts as a similarity measure between images. If the two images are exactly matched, then $\boldsymbol{p}_i = (\delta_{ij})_{1 \le j \le n}$, $i = 1, 2, \ldots, n$. Since $\boldsymbol{p}_i$s are degenerate distributions, by Proposition 2, the Jensen–Rényi divergence is maximized for a fixed α and $\boldsymbol{\omega}$. Fig. 6(1)–(2) shows two brain MRT images in which the misalignment is a Euclidean rotation. The conditional probability distributions $\{\boldsymbol{p}_i\}$ are crisp, as in Fig. 6(3), when the two images are aligned, and dispersed, as in Fig. 6(4), when they are not matched.

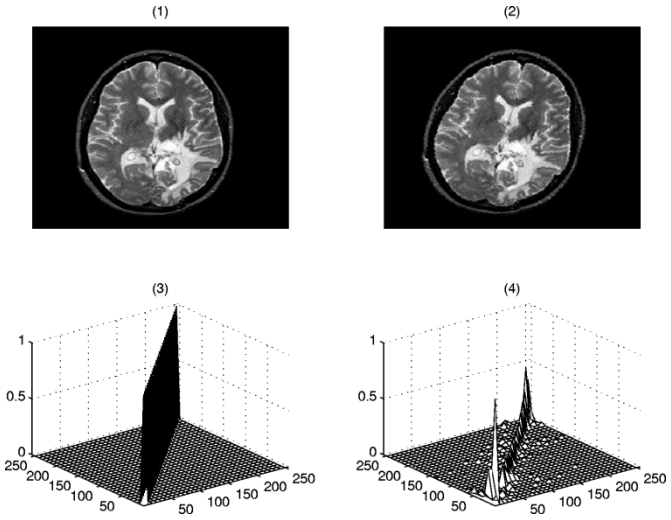

Fig. 6. Conditional probability distributions.

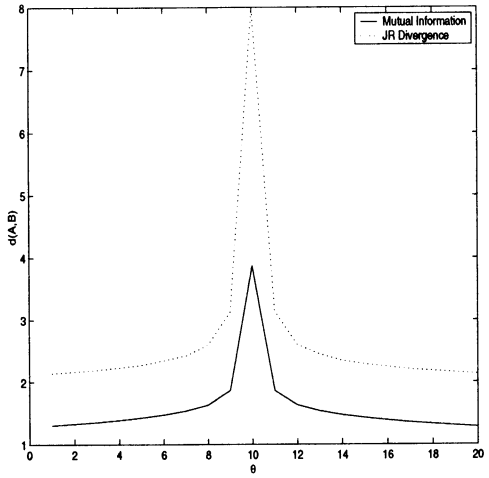

Fig. 7. Mutual information versus Jensen–Rényi divergence of uniform weights.

It is worth noting that the maximization of the Jensen–Rényi divergence holds for any α and $\boldsymbol{\omega}$ such that $0 \leq \alpha \leq 1$ and $\omega_i \geq 0$, $\sum_i \omega_i = 1$. If we take $\alpha = 1$ and $\omega_i = P(X = x_i)$, then by Proposition 1, the Jensen–Rényi divergence is exactly the Shannon mutual information. Indeed, the Jensen–Rényi divergence induced similarity measure provides a more general framework for the image registration problem.

If the two images f_1 and $\mathcal{T}_{(\ell,\theta,\gamma)} f_2$ are matched, the Jensen–Rényi divergence is maximized for any valid weight. Assigning $\omega_i = P(X = x_i)$ is not always a good choice. Fig. 7 shows the registration results of the two brain images in Fig. 6 using the mutual information and the Jensen–Rényi divergence of $\alpha = 1$ and uniform weights. The peak at the matching point generated by the Jensen–Rényi divergence is clearly much higher than the peak by the mutual information. $\omega_i = P(X = x_i)$ gives the background pixels the largest weights. In the presence of noise, the matching in background is corrupted. Mutual information may fail to identify the registration point. This phenomenon is demonstrated in Fig. 8. The following proposition establishes the optimality of the

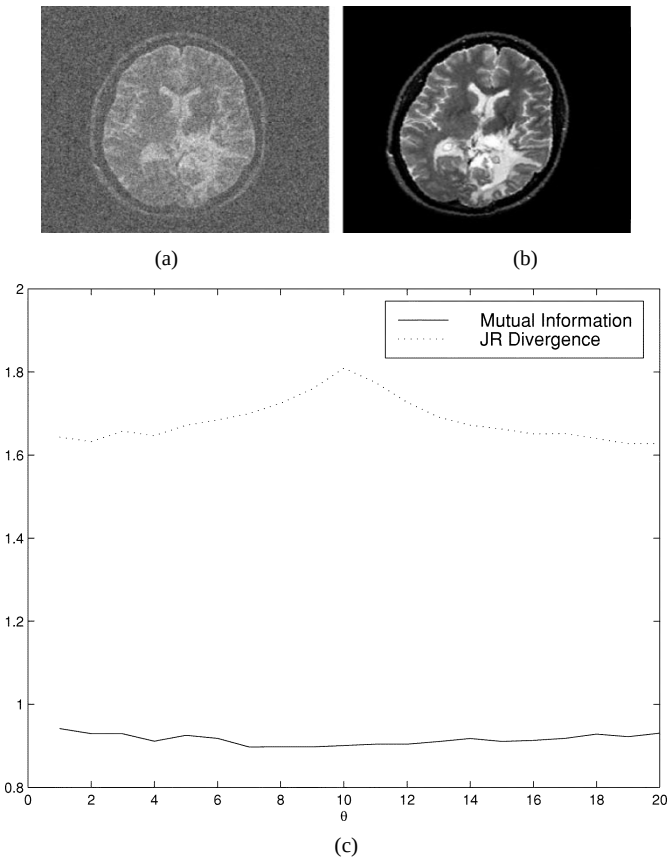

Fig. 8. Registration result in the presence of the noise. SNR= 1.92 dB. For the Jensen–Rényi divergence, $\alpha = 1$ and $\omega_i = 1/n$ is used. (a) Image A. (b) Image B. (c) $d(A, B)$.

uniform weights for image registration in the context of the Jensen–Rényi divergence.

Proposition 4: Let $\boldsymbol{\beta}$ be a uniform weight defined as $\beta_i = 1/n$, $i = 1, 2, \ldots, n$, and let $\boldsymbol{\omega}$ be any vector such that $\omega_i \geq 0$, $\sum_{i=1}^{n} \omega_i = 1$. If the misalignment between f_1 and f_2 can be modeled by a spatial transformation $\mathcal{T}^*$, then for all $\alpha \in [0, 1]$, the following inequality holds:

$$\mathrm{JR}_\alpha^{\boldsymbol{\beta}}(\boldsymbol{p}_1(f_1, \mathcal{T}^* f_2), \ldots, \boldsymbol{p}_n(f_1, \mathcal{T}^* f_2)) \geq JR_\alpha^{\boldsymbol{\omega}}(\boldsymbol{p}_1(f_1, \mathcal{T}^* f_2), \ldots, \boldsymbol{p}_n(f_1, \mathcal{T}^* f_2)).$$

Proof: $\boldsymbol{p}_i = \Delta_i$, $i = 1, 2, \ldots, n$ when f_1 and f_2 are aligned by the spatial transformation $\mathcal{T}^*$; then, $JR_\alpha^{\boldsymbol{\omega}}(\cdot)$ becomes

$$\mathrm{JR}_\alpha^{\boldsymbol{\omega}}(\boldsymbol{p}_1(f_1, \mathcal{T}^* f_2), \ldots, \boldsymbol{p}_n(f_1, \mathcal{T}^* f_2)) = R_\alpha\left(\sum_{i=1}^{n} \omega_i \Delta_i\right) = R_\alpha(\boldsymbol{\omega}).$$

Since $\beta \prec \boldsymbol{\omega}$ [31] and $R_\alpha(\cdot)$ is Schur-concave, we obtain $R_\alpha(\beta) \geq R_\alpha(\boldsymbol{\omega})$. This completes the proof. ■

After assigning uniform weights to the various distributions in the Jensen–Rényi divergence, a free parameter α, which is directly related to the measurement sensitivity, remains to be selected. In the image registration problem, one desires a sharp and distinguishable peak at the matching point. The sharpness of the Jensen–Rényi divergence can be characterized by the maximal value as well as the width of the peak. The sharpest peak is clearly a Dirac function. The following proposition establishes

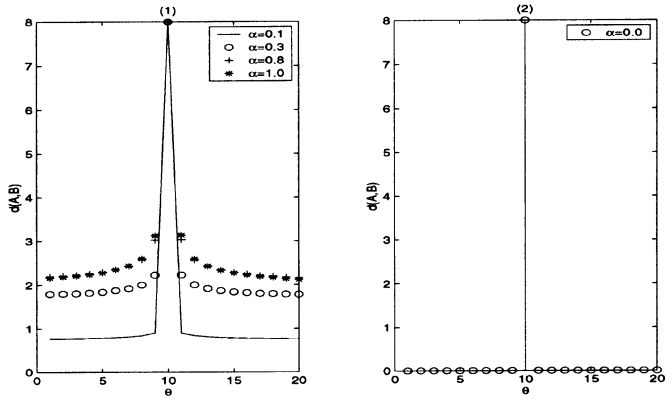

Fig. 9. Effect of the order α in image registration.

that the maximal value of the Jensen–Rényi divergence is independent of α if the two images are aligned, and $\alpha = 0$ yields the sharpest peak, which is a Dirac function.

Proposition 5: Let β be a uniform weight vector. If the misalignment between f_1 and f_2 can be modeled by a spatial transformation $\mathcal{T}^*$, then for all $\alpha \in [0, 1]$, we have

$$\mathrm{JR}_\alpha^\beta(\boldsymbol{p}_1(f_1, \mathcal{T}^* f_2), \ldots, \boldsymbol{p}_n(f_1, \mathcal{T}^* f_2)) = \log n. \tag{13}$$

In case of $\alpha = 0$

$$\mathrm{JR}_\alpha^\beta(\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n) = 0$$

for any probability distribution $\boldsymbol{p}_i$ such that $p_{ij} > 0$, $i, j = 1, 2, \ldots, n$ and

$$\mathrm{JR}_\alpha^\beta(\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n) = \log(n)$$

if and only if $\boldsymbol{p}_i = \Delta_i$, $i = 1, 2, \ldots, n$.

Proof: See the Appendix. ■

As an example, Fig. 9(a) demonstrates the registration results of the two brain images in Fig. 6 with the choice of different α. In this case, $\alpha = 0$ is the best choice and would generate a Dirac function with a peak at the matching point, as illustrated in Fig. 9(b).

If there exists local variation between f_1 and f_2, or if the registration of the two images is in the presence of noise, then an exact alignment $\mathcal{T}^*$ may not be found. The conditional probability distribution $\boldsymbol{p}_i(f_1, \mathcal{T}^* f_2)$ is no longer a degenerate distribution in this case. The following proposition establishes that taking $\alpha = 1$ would provide a higher peak than any other choice of α for the nonideal alignment.

Proposition 6: Let $\boldsymbol{p}_i = \Delta_i + \delta\boldsymbol{p}_i$, $i = 1, 2, \ldots, n$, where $\delta\boldsymbol{p}_i = (\delta p_{ij})_{1 \le j \le n}$ is a real distortion vector such that $p_{ij} \ge 0$, $\sum_{j=1}^n \delta p_{ij} = 0$, and $\sum_{i=1}^n \delta p_{ij} = 0$. Let $\boldsymbol{\omega}$ be a weight vector, and denote $\mathrm{JR}^{\boldsymbol{\omega}}(\cdot)$ as the Jensen–Rényi divergence with $\alpha = 1$. Then, for all $\alpha \in (0, 1)$, we have

$$\mathrm{JR}^{\boldsymbol{\omega}}(\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n) \ge JR_\alpha^{\boldsymbol{\omega}}(\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n). \tag{14}$$

Proof: Observe that for any probability distribution $\boldsymbol{p}$, $R_\alpha(\boldsymbol{p}) \ge H(\boldsymbol{p})$, $\forall\, \alpha \in (0, 1)$, and then

$$\sum_{i=1}^n \omega_i H(\boldsymbol{p}_i) \le \sum_{i=1}^n \omega_i R_\alpha(\boldsymbol{p}_i), \qquad \forall\, \alpha \in (0, 1). \tag{15}$$

Since $\sum_{j=1}^n \delta p_{ij} = 0$, $\sum_{i=1}^n \delta p_{ij} = 0$ and the Rényi entropy of $\alpha = 1$ is exactly the Shannon entropy, (14) is equivalent to the inequality (15). This completes the proof. ■

It is worth pointing out that the Jensen–Rényi divergence is not equivalent to mutual information by setting $\alpha = 1$. The equivalence is hold only if $\alpha = 1$ and $\omega_i = P(X = x_i)$.

A. Discussion

Parameter α basically plays a role of scaling factor to adjust registration peaks, and the location of registration point is independent of α. In real-world applications, there is a tradeoff between optimality and practicality in choosing α. If one can model the misalignment between f_1 and f_2 completely and accurately, $\alpha = 0$ would correspond to the best choice since it generates a Dirac function at the matching point. It is, however, also the least robust selection, as it tends to make all the $\boldsymbol{p}_i$s the same as the uniform distribution. If $\boldsymbol{p}_i$ is not degenerate distribution and $p_{ij} > 0$, then the Jensen–Rényi divergence would be zero for the whole transformation parameter space as in case where the adapted transformation group cannot accurately model the relationship between f_1 and f_2. On the other hand, $\alpha = 1$ is the most robust choice, in spite of also resulting in the least sharp peak. The choice of α therefore depends largely on the accuracy of the invoked model and on the specific application as well as the available computational resource.

We further showed that $\omega_i = 1/n$ is optimal; thus, the best choice for nonideal image registration in the context of the Jensen–Rényi divergence is $\{\alpha = 1, \omega = 1/n\}$, in comparison with mutual information based methods, in which the parameters are set to $\{\alpha = 1, \omega = P(X = x_i)\}$. Fig. 8 demonstrates the registration results by mutual information and by the Jensen–Rényi divergence, in the presence of the noise. SNR $= 1.92$ dB. For the Jensen–Rényi divergence, $\alpha = 1$ and $\omega_i = 1/n$ is used.

Computational complexity for registration with Jensen–Rényi divergence would eventually depends on the computation of $n + 1$ Rényi entropy. Rényi entropy is computed using the histogram method, which requires preparing a conditional pixel intensity histogram. When using all 256 gray levels of the original images, the resulting conditional histogram represents a 256×256 discrete matrix. In this case, $n = 256$. An average distribution was then prepared by summing the weighted conditional histogram entries along the axis corresponding to the image intensity index of f_1. Jensen–Rényi divergence eventually calculates the difference between the Rényi entropy of the average distribution and the sum of n weighted Rényi entropy of individual distributions. For the registration of two brain MRT images in which the misalignment is a Euclidean rotation, as shown in Fig. 6(a)–(b), the Matlab simulation takes less than 2 min on a Pentium III 700 MHz machine with 128 MB of memory.

V. Numerical Experiments: ISAR Image Registration

Generating an ISAR image by using stepped frequency waveform [26] can be understood as a process of estimating the target's 2-D reflectivity density function $\rho(x, y)$ from data collected in the frequency space. Suppose a stepped frequency burst consists of M pulses in which the transmitted frequency linearly increases from ω_0 to $\omega_0 + (M - 1)\Delta\omega$, where ω_0 is

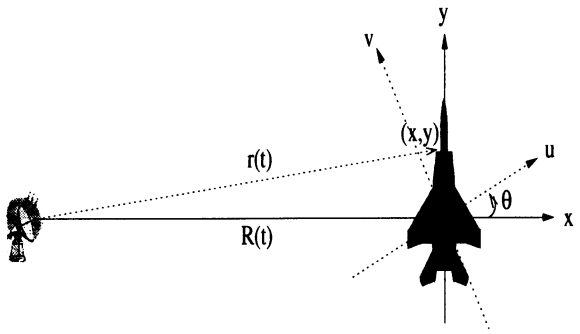

Fig. 10. ISAR geometry of a moving target.

the base frequency in radians per second, and $\Delta\omega$ is the step frequency. Let the mth transmitted pulse $s_m(t)$ be a pulse of duration T_p and expressed in complex form as

$$s_m(t)=e^{j\omega_m t}W\left(\frac{t-mT_p}{T_p}\right), \quad m=0,\ldots,M-1 \tag{16}$$

where $\omega_m = \omega_0 + m\Delta\omega$, and

$$W(t)=\begin{cases}1, & 0\le t<1\\ 0, & \text{otherwise.}\end{cases} \tag{17}$$

Define $\omega(t) = \omega_0 + (m-1)\Delta\omega, mT_p \le t < (m+1)T_p$, $m = 0, 1, \ldots, M-1$. Under uniform illumination, the reflected signal from the target differential area $dx \times dy$ at the target coordinate (x, y) is

$$h(x, y, t) = A\rho(x, y)e^{j\omega(t)(t-2r(t)/c)}\,dx\,dy \tag{18}$$

where $0 \le t < (M-1)T_p$, and A is a constant attenuation factor, which we can set to 1 without a loss of generality. The distance between the radar antenna and the target reflection point located at (x, y) is denoted by $r(t)$. We obtain the expression of the received signal for $0 \le t < (M-1)T_p$ by integrating reflections from all the point scatterers in the target

$$\begin{aligned}\tilde{g}(t) &= \iint_{\mathbb{R}^2} h(x, y, t)\,dx\,dy\\ &= \iint_{\mathbb{R}^2} \rho(x, y)e^{j\omega(t)(t-2r(t)/c)}\,dx\,dy.\end{aligned} \tag{19}$$

After quadrature demodulation, we obtain

$$g(t) = \iint_{\mathbb{R}^2} \rho(x, y)e^{-j2\omega(t)r(t)/c}\,dx\,dy. \tag{20}$$

It can be observed from Fig. 10 that for target dimensions that are relatively smaller than the target range R, the distance $r(\cdot)$ from the radar antenna to target reflection point located at (x, y) is

$$r(t) \approx R(t) + x\cos\theta(t) - y\sin\theta(t). \tag{21}$$

Inserting (21) into (20), we deduce the baseband signal in terms of target coordinate (x, y) and rotation angle θ

$$g(t) = e^{-j2\omega(t)(R(t)/c)} \iint_{\mathbb{R}^2} \rho(x, y)e^{-j(x\omega_x(t)-y\omega_y(t))}\,dx\,dy \tag{22}$$

where

$$\omega_x(t) = \frac{2\omega(t)}{c}\cos\theta(t) \tag{23}$$

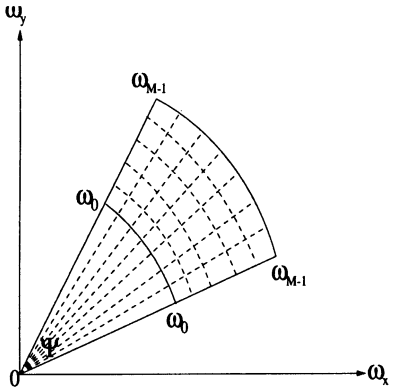

Fig. 11. Polar-formatted data in spatial frequency space.

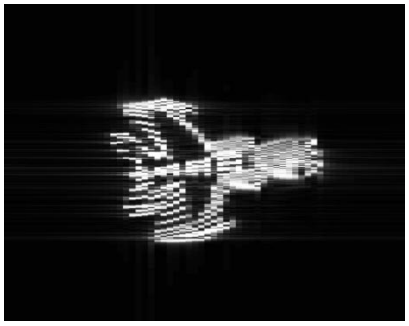

Fig. 12. ISAR image of moving target reconstructed by the discrete Fourier transformation.

and

$$\omega_y(t) = \frac{2\omega(t)}{c}\sin\theta(t) \tag{24}$$

are spatial frequency quantities defined at frequency $\omega(t)$ and target rotation angle $\theta(t)$. The phase term $e^{-j2\omega(t)R(t)/c}$ is related to the target translational motion only and can be compensated by traditional translational motion-compensation methods.

By sampling $e^{j2\omega(t)R(t)/c}g(t)$ at $t_m = (m+(1/2))T_p$, $m = 0, 1, \ldots, M-1$, we obtain the data collected in the frequency space $G(m)$ as

$$G(m) = \iint_{\mathbb{R}^2} \rho(x, y)e^{-j(x\omega_x(t_m)-y\omega_y(t_m))}\,dx\,dy \tag{25}$$

where $m = 0, 1, \ldots, M-1$.

To form a radar image, N bursts of received signal are sampled and organized burst by burst into a $M \times N$ 2-D array, which is shown in Fig. 11. This sample matrix is not uniformly spaced in the spatial frequency; instead, it is polar-formatted data. The discrete Fourier transform processing of the polar-formatted data would result in blurring at the edges of the target reflectivity image. Fig. 12 is a synthetic ISAR image of a MIG-25 aircraft [32]. The radar is assumed to be operating at 9 GHz and transmits a stepped-frequency waveform. Each burst consists of 64 narrowband pulses stepped in frequency from pulse to pulse by a fixed frequency step of 8 MHz. The pulse repetition frequency is 15 KHz. Basic motion-compensation processing has been applied to the data. A total of 512 bursts of a received signal are taken to reconstruct the image of this aircraft, which corresponds to 2.18 s integration time. As we can see, the

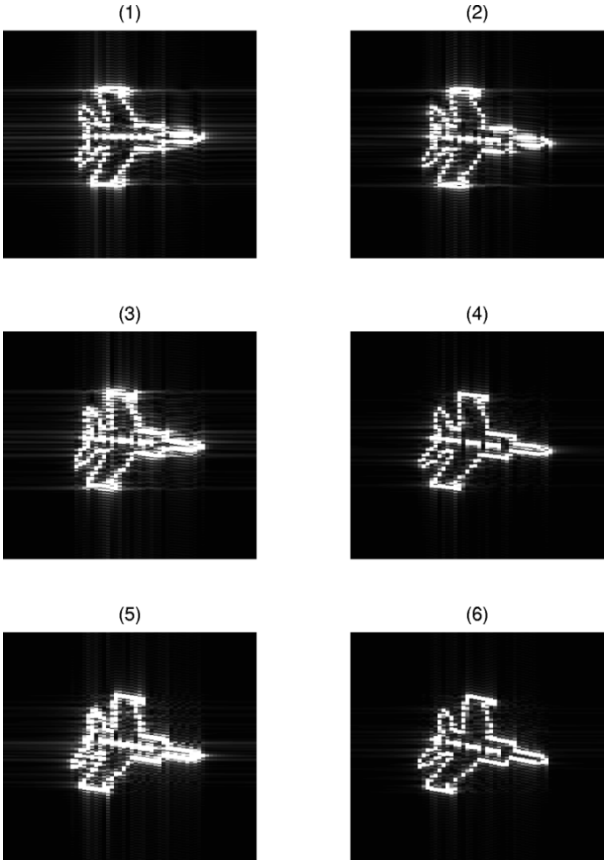

Fig. 13. Trajectory of a sequence of MIG-25 image frames.

resulting image is defocused due to the target rotation. In fact, the defocused image in Fig. 12 is formed by overlapping a series of MIG-25s at different viewing angles. By replacing the Fourier transform with the time-varying spectral analysis techniques [32], [33], we can take a sequence of snapshots of the target during the 2.18 s of integration time. Fig. 13(1)–(6) shows the trajectory of the MIG-25 with six image frames taken at $t =$ 0.1280, 0.4693, 0.8107, 1.1520, 1.4933, and 1.8347 s, respectively.

Image registration can be applied to estimate the target motion from this sequence of images. For the synthetic ISAR images shown in Fig. 13, we search for the rotation angles $\{\theta_i\}_{i=1}^N$ between a sequence of image frames $\{I_i\}_{i=0}^N$ observed in a time interval $[0, T]$. By (12), θ_i is given by

$$\theta_i^* = \arg\max_{\theta_i} \mathrm{JR}_\alpha^{\boldsymbol{\omega}}(\boldsymbol{p}_1(I_{i-1}, \mathcal{T}_{\theta_i} I_i), \ldots, \boldsymbol{p}_n(I_{i-1}, \mathcal{T}_{\theta_i} I_i)).$$

Fig. 14 shows the rotation angles $\{\theta_i\}_{i=1}^N$ obtained by registering the six consecutive MIG-25 image frames. As can already be seen in the figures, uniform weights produce the sharpest peak.

By interpolating $\{\theta_i\}_{i=1}^N$, we obtain a trajectory of the MIG-25 rotational motion during the imaging time as shown in the bottom right-hand plot of Fig. 14. The latter is particularly important since it may be subsequently used in polar reformatting [26] and resampling the received signal into rectangular format. This results in a focused image of the MIG-25 based on all the received signals in the time interval $[0, 2.18\,\mathrm{s}]$, as demonstrated in Fig. 15.

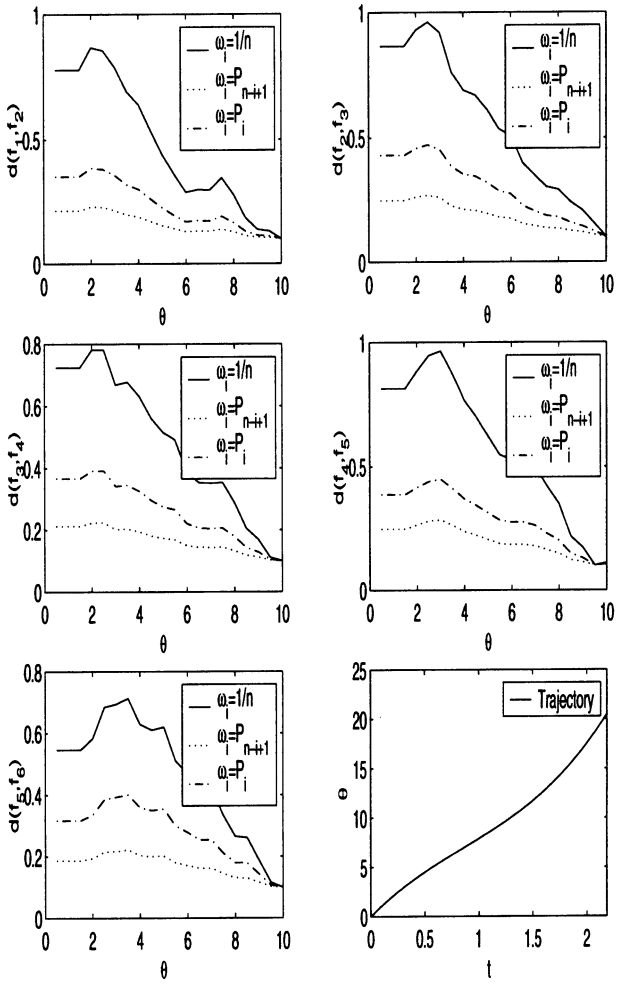

Fig. 14. Image registration of a MIG-25 trajectory.

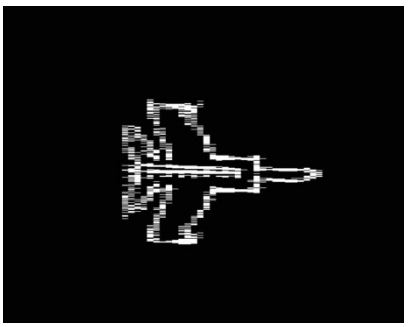

Fig. 15. Reconstructed MIG-25 by polar reformatting.

VI. Conclusions

In this paper, we proposed a new information divergence measure, i.e., *Jensen–Rényi* divergence, and analyzed its properties. We studied the relationship between mutual information and Jensen–Rényi divergence and concluded that mutual information can be enclosed in the framework of the Jensen–Rényi divergence, and this gives us a starting point for image registration problem. We further showed the registration performance difference between mutual information (equivalent to JR with $\alpha = 1$ and $\omega_i = P(x_i)$) and Jensen–Rényi divergence (with $\alpha = 1$ and $\omega_i = 1/n$) and point out that assigning $\omega_i = P(x_i)$

is not a good choice since it tends to emphasize background in the process of matching. To demonstrate this phenomenon, a breaking point of mutual information registration under noisy situation (SNR < 2 dB) is presented, while at the same time, Jensen–Rényi divergence is still able to identify the registration point. We also showed the scaling effect of order α in the Jensen–Rényi divergence, which would adjust the peak of registration point. As the key application of this paper, experiments of ISAR data are presented for rotational motion compensation. Our approach applies Jensen–Rényi divergence to measure the spatial dependence between consecutive ISAR image frames and estimates the target trajectory during the imaging time, which enables polar reformatting to resample the received signal into a rectangular format. Simulation results demonstrate that the proposed method effectively focuses the target image.

APPENDIX

Proof of Proposition 1

Denote $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ and $\mathcal{Y} = \{y_1, y_2, \ldots, y_n\}$. Let X, Y be two random variables taking values in $\mathcal{X}$ and $\mathcal{Y}$. Recall that the mutual information between X and Y is given by [34]

$$I(X;Y) = H(Y) - H(Y|X) \tag{26}$$

where $H(Y)$ is the Shannon entropy of Y, and $H(Y|X)$ is the conditional Shannon entropy of Y given X.

Instead of using Shannon entropy in (26), the mutual information can be generalized using Rényi entropy. Therefore, the α-mutual information can be defined as

$$I_\alpha(X;Y) = R_\alpha(Y) - R_\alpha(Y|X), \qquad \alpha \in (0, 1)$$

where R_α is the Rényi entropy of order $\alpha \in (0, 1)$.

Denote by $P(x_i) = \omega_i$, $P(Y = y_j|X = x_i) = p_{ij}$, and $P(Y = y_j) = q_j$; then, it is easy to check that

$$R_\alpha(Y) - R_\alpha(Y|X) = JR_\alpha^{\boldsymbol{\omega}}(\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n) \tag{27}$$

where $\boldsymbol{p}_i = (p_{ij})_{1 \le j \le k}$, for all $i = 1, \ldots, n$.

For fixed ω_i, the mutual information is a convex function of p_{ij} [34]; then, it can be verified that the α-mutual information is also a convex function of p_{ij}, leading to the Jensen–Rényi divergence a convex function of $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$. ■

Proof of Proposition 5

Using Proposition 4, we have

$$\begin{aligned} \mathrm{JR}_\alpha^{\boldsymbol{\beta}}(\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n) &= R_\alpha\left(\sum_{i=1}^n \beta_i \Delta_i\right) \\ &= \frac{1}{1-\alpha} \log \sum_{i=1}^n \left(\frac{1}{n}\right)^\alpha \\ &= \log(n). \end{aligned}$$

For $\alpha = 0$

$$\begin{aligned} R_\alpha(\boldsymbol{p}_i) &= \frac{1}{1-\alpha} \log \sum_{i=1}^n (p_{ij})^\alpha \\ &= \log n, \qquad \forall p_{ij} > 0, \sum_{j=1}^n p_{ij} = 1 \end{aligned}$$

we obtain

$$\mathrm{JR}_\alpha^{\boldsymbol{\beta}}(\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n) = R_\alpha\left(\sum_{i=1}^n \frac{1}{n} \boldsymbol{p}_i\right) - \sum_{i=1}^n \frac{1}{n} R_\alpha(\boldsymbol{p}_i) = 0.$$

If $\alpha = 0$ and $\boldsymbol{p}_i = \Delta_i$, then Proposition 4 leads to

$$\mathrm{JR}_\alpha^{\boldsymbol{\beta}}(\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n) = \log(n).$$

This concludes the proof for the proposition. ■

REFERENCES

[1] R. Kasturi and R. C. Jain, *Computer Vision: Principles*. Los Alamitos, CA: IEEE Comput. Soc., 1991.

[2] B. K. P. Horn, *Robot Vision*. Cambridge, MA: MIT Press, 1989.

[3] J. R. Jensen, *Introductory Digital Image Processing: A Remote Sensing Perspective*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1996.

[4] I. L. Thomas, V. M. Benning, and N. P. Ching, *Classification of Remotely Sensed Images*. Bristol, U.K.: Adam Hilger, 1986.

[5] M. R. Stytz, S. Frieder, and O. Frieder, "Three dimensional medical imaging: Algorithms and computer systems," *ACM Comput. Surveys*, vol. 23, no. 4, pp. 421–424, 1991.

[6] P. A. Van den Elsen, E.-J. D. Pol, and M. A. Viergever, "Medical image matching—A review with classification," *IEEE Eng. Med. Biol. Mag.*, vol. 12, pp. 26–39, Jan. 1993.

[7] A. K. Gabriel and R. M. Goldstein, "Crossed orbit interferometry: Theory and experimental results from SIR-B," *Int. J. Remote Sens.*, vol. 9, pp. 857–872, 1988.

[8] F. K. Li and R. M. Goldstein, "Studies of multibaseline spaceborne interferometric synthetic aperture radars," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, pp. 88–97, Feb. 1990.

[9] Q. Lin, J. F. Vesecky, and H. A. Zebker, "New approaches in interferometric SAR data processing," *IEEE Trans. Geosci. Remote Sensing*, vol. 30, pp. 560–567, June 1992.

[10] G. Fornaro and G. Franceschetti, "Image registration in interferometric SAR processing," *Proc. Inst. Elect. Eng., Radar, Sonar Navigat.*, vol. 142, no. 6, pp. 313–320, 1995.

[11] L. Brown, "A survey of image registration techniques," *ACM Computi. Surveys*, vol. 24, no. 4, pp. 325–376, 1992.

[12] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*. Orlando, FL: Academic, 1982.

[13] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," in *Proc. IEEE Int. Conf. Cybern. Soc.*, 1975, pp. 163–165.

[14] E. De Castro and C. Morandi, "Registration of translated and rotated images using finite Fourier transforms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 700–703, 1987.

[15] A. Goshtasby, "Image registration by local approximation," *Image Vision Comput.*, vol. 6, no. 4, pp. 255–261, 1988.

[16] C. Broit, "Optimal registration of deformed images," Ph.D. dissertation, Univ. Penna., Philadelphia, PA, 1981.

[17] R. Bajscy and S. Kovacic, "Multiresolution elastic matching," *Comput. Vision Graph. Image Process.*, vol. 46, pp. 1–21, 1989.

[18] D. J. Burr, "A dynamic model for image registration," *Comput. Graphics Image Process.*, vol. 15, pp. 102–112, 1981.

[19] R. P. Woods, J. C. Mazziotta, and S. R. Cherry, "MRI-PET registration with automated algorithm," *J. Comput. Assist. Tomogr.*, vol. 17, no. 4, pp. 536–546, 1993.

[20] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, 1997.

[21] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, pp. 187–198, Mar. 1997.

[22] A. O. Hero, III and O. J. J. Michel, "Asymptotic theory of greedy approximations to minimal k-point random graphs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1921–1938, June 1999.

[23] J. W. Fisher, III and A. S. Willsky, "Information theoretic feature extraction for ATR," in *Proc. 34th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, 1999.

[24] A. Tsai, J. W. Fisher, III, C. Wible, W. W. Wells, III, J. Kim, and A. S. Willsky, "Analysis of functional MRI data using mutual information," in *Proc. Second Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, Cambridge, U.K., 1999.

[25] A. Rényi, "On measures of entropy and information," in *Selected Papers of Alfréd Rényi*, 1976, vol. 2, pp. 525–580.

[26] D. R. Wehner, *High Resolution Radar*, 2nd ed. Norwood, MA: Artech House, 1995.

[27] D. C. Munson and R. L. Visentin, "A signal processing view of strip-mapping synthetic aperture radar," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 2131–2147, Dec. 1989.
[28] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.
[29] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Inform. Theory*, vol. 37, pp. 145–151, Jan. 1991.
[30] T. Bisztriczky, P. McMullen, R. Schneider, and A. W. Weiss, *Polytopes: Abstract, Convex, and Computational*. Dordrecht, The Netherlands: Kluwer, 1994.
[31] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.
[32] V. C. Chen and S. Qian, "Joint time frequency transform for radar range-doppler imaging," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, pp. 486–499, Feb. 1998.
[33] Y. He, A. Ben Hamza, H. Krim, and V. C. Chen, "An information theoretic measure for ISAR imagery focusing," in *Proc. SPIE*, vol. 4116, 2000, pp. 463–471.
[34] G. Gallager, *Information Theory and Reliable Communications*. London, U.K.: Wiley, 1968.

Yun He received the B.S. and M.S. degrees from Beijing University of Aeronautics and Astronautics, Beijing, China, in 1995 and 1998, and the Ph.D. degree from North Carolina State University, Raleigh, in 2001, all in electrical engineering.

From May 1998 to July 2001, he was a Research Assistant with the Center for Advanced Computing and Communications, North Carolina State University. He then joined the Mixed Signal IC Design Group at Tality Corporation, Cary, NC. His research interests are in wavelet theory and applications and real-time digital signal processing.

A. Ben Hamza received the B.S. and M.S. degrees in applied mathematics. From March 2000 to February 2001, he was a Research Associate with the Department of Electrical and Computer Engineering at North Carolina State University, Raleigh, where he is currently pursuing the Ph.D. degree.

His research interests include nonlinear probabilistic and variational filtering, information-theoretic measures, and computer vision.

Hamid Krim (M'80–SM'98) received the B.S., M.S., and Ph.D. degrees in electrical engineering.

As a Member of Technical Staff at AT&T Bell Labs, he has worked in the areas of telephony and digital communication systems/subsystems. Following an NSF postdoctoral fellowship at Foreign Centers of Excellence, LSS/University of Orsay, Paris, France, he became a Research Scientist at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, performing and supervising research. He is presently with the Electrical and Computer Engineering Department, North Carolina State University, Raleigh. His research interests are in statistical signal and image analysis and mathematical modeling with a keen emphasis on applied problems.