Optics and Lasers in Engineering 48 (2010) 1213–1223

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng

Medical image registration using stochastic optimization

Waleed Mohamed, A. Ben Hamza *

Concordia Institute for Information Systems Engineering, Concordia University, 1515 Ste-Catherine Street West, Montréal, QC, Canada EV7.631

ARTICLE INFO

Article history:
Received 28 January 2010
Received in revised form
28 March 2010
Accepted 27 June 2010
Available online 17 July 2010

Keywords:
Image registration
Tsallis entropy
Divergence measure
Stochastic optimization

ABSTRACT

We propose an image registration method by maximizing a Tsallis entropy-based divergence using a modified simultaneous perturbation stochastic approximation algorithm. Due to its convexity property, this divergence measure attains its maximum value when the conditional intensity probabilities between the reference image and the transformed target image are degenerate distributions. Experimental results are provided to demonstrate the registration accuracy of the proposed approach in comparison to existing entropic image alignment techniques.

1. Introduction

Image registration or alignment refers to the process of aligning images so that their details overlap accurately [1,2]. Images are usually registered for the purpose of combining or comparing them, enabling the fusion of information in the images. Roughly speaking, the image alignment problem may be formulated as a two-step process: the first step is to define a dissimilarity measure that quantifies the quality of spatial alignment between the reference image and the spatially transformed target image, and the second step is to develop an efficient optimization algorithm for maximizing this dissimilarity measure in order to find the optimal transformation parameters. Recently, much attention has been paid to the image registration problem due to its importance in a variety of tasks including data fusion, navigation, motion detection, and clinical studies [1,2]. A wide range of image registration techniques have been developed for many different types of applications and data, such as mean squared alignment, correlation registration, moment invariant matching, and entropic alignment [3–6]. The latter will be the focus of this paper. Inspired by the successful application of the mutual information measure [3,4], and looking to address its limitations in often difficult imagery, we proposed in [7] an information-theoretic approach to ISAR image registration by estimating the target motion during the imaging time, and it was accomplished using the Jensen–Rényi divergence. This generalized entropic measure enjoys appealing mathematical properties affording a great flexibility in a number of applications [8–11].

In recent years, there has been a concerted research effort in statistical physics to explore the properties of Tsallis entropy, leading to a statistical mechanics that satisfies many of the properties of the standard theory [12]. In [13], a Tsallis entropy-based image mutual information approach, combined with a stochastic optimization algorithm, was proposed leading to accurate image registration results compared to the classical mutual information [3,4].

In this paper, we propose an entropic image alignment approach by maximizing the Jensen–Tsallis divergence [14] using a simultaneous perturbation stochastic approximation-based algorithm [15]. The main contributions in this paper may be summarized as follows: (i) explore the use of the Jensen–Tsallis divergence as an alignment measure, (ii) develop an efficient optimization algorithm to maximize this divergence measure, and (iii) perform an experimental comparative study of the proposed approach with existing entropic image registration methods.

The rest of this paper is organized as follows. The next section is devoted to the problem formulation, followed by a theoretical analysis of the Jensen–Tsallis divergence and a derivation of its upper bound. Then, we develop a modified simultaneous perturbation stochastic approximation algorithm to maximize the divergence measure. In Section 3, we describe the proposed image alignment method and discuss its most important algorithmic steps in more details. In Section 4, we provide experimental results to show the effectiveness and the registration accuracy of the proposed approach. And finally, we conclude and point out possible future work directions in Section 5.

* Corresponding author.
E-mail address: hamza@ciise.concordia.ca (A. Ben Hamza).

0143-8166/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.optlaseng.2010.06.011

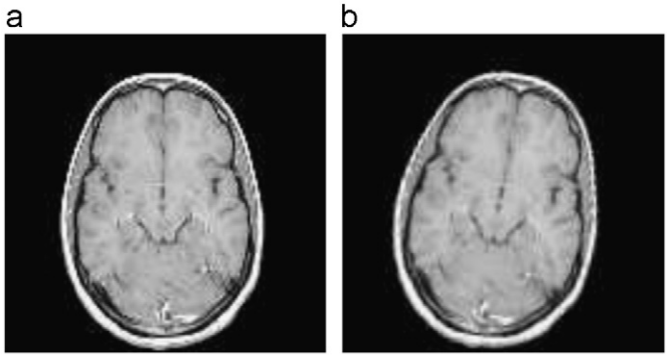


Fig. 1. (a) Reference image I_1 and (b) target image I_2.

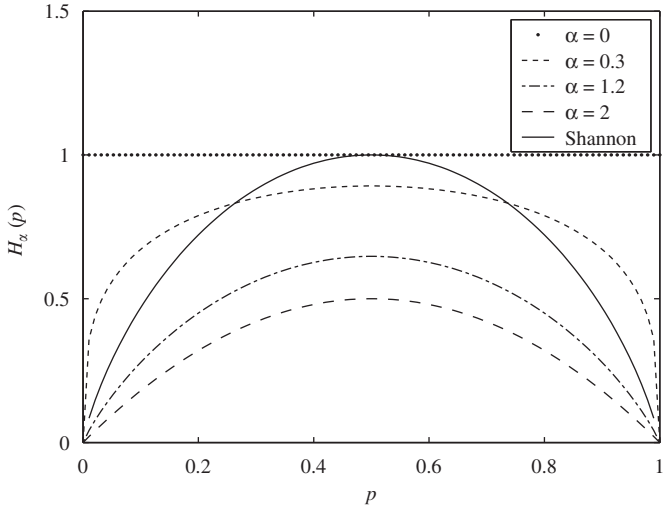


Fig. 2. Tsallis entropy $H_\alpha(\boldsymbol{p})$ of a Bernoulli distribution $\boldsymbol{p}=(p,1-p)$ for different values of α.

2. Problem formulation

In the continuous domain, an image is defined as a real-valued function $I:\Omega\rightarrow\mathbb{R}$, and Ω is a nonempty, bounded, open set in $\mathbb{R}^2$ (usually Ω is a rectangle in $\mathbb{R}^2$). Throughout, we denote by $\boldsymbol{x}=(x_1,x_2)$ a pixel location in Ω. Given two misaligned images, the reference image I_1 and the target image I_2 as depicted in Fig. 1, the image alignment or registration problem may be formulated as an optimization problem

$$\ell^* = \arg\max_{\ell} D(I_1(\boldsymbol{x}), I_2(\Phi_\ell(\boldsymbol{x}))), \tag{1}$$

where $D(\cdot,\cdot)$ is a dissimilarity measure that quantifies the discrepancy between the reference image and the transformed target image; and $\Phi_\ell:\Omega\leftarrow\Omega$ is a spatial transformation parameterized by a parameter vector ℓ. An example of such a mapping is a Euclidean transformation with a parameter vector $\ell=(\boldsymbol{t},\theta,\boldsymbol{s})$, where $\boldsymbol{t}=(t_x,t_y)$ is a translational parameter vector, θ is a rotational parameter, and $\boldsymbol{s}=(s_x,s_y)$ is a scaling parameter vector.

The goal of image registration is to align the target image to the reference image by maximizing the dissimilarity measure $D(I_1(\boldsymbol{x}),I_2(\Phi_\ell(\boldsymbol{x})))$ using an optimization scheme in order to find the optimal spatial transformation parameters. Note that since the image pixel values are integers, a bilinear interpolation may be used to determine the values of $I_2(\Phi_\ell(\boldsymbol{x}))$ when $\Phi_\ell(\boldsymbol{x})$ is not an integer. In this paper, we use the Jensen–Tsallis divergence as a dissimilarity measure [14], and a modified simultaneous perturbation stochastic approximation (SPSA) approach as an optimization algorithm [15].

2.1. Jensen–Tsallis divergence

Let $X=\{x_1,x_2,\ldots,x_k\}$ be a finite set with a probability distribution $\boldsymbol{p}=(p_1,p_2,\ldots,p_k)$ where $k>1$. Shannon's entropy is defined as $H(\boldsymbol{p})=-\sum_{j=1}^{k} p_j\log(p_j)$, and it is a measure of uncertainty,

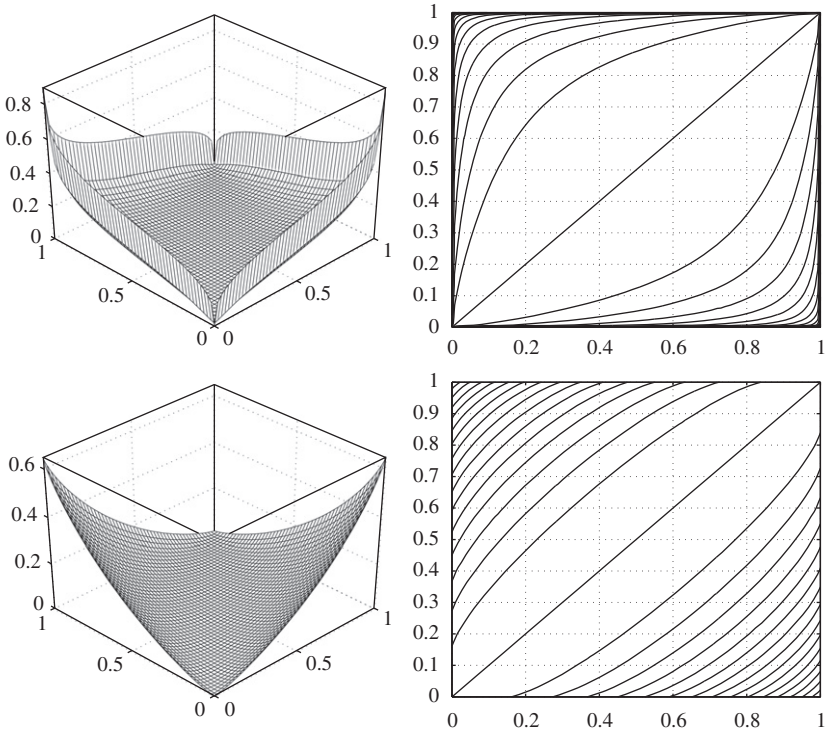


Fig. 3. Surface/contour plots of Jensen–Tsallis divergence between two Bernoulli distributions $\boldsymbol{p}=(p,1-p)$ and $\boldsymbol{q}=(q,1-q)$, and with equal weights $\omega_1=\omega_2=1/2$: first row: $\alpha=0.3$ and second row: $\alpha=1.2$.

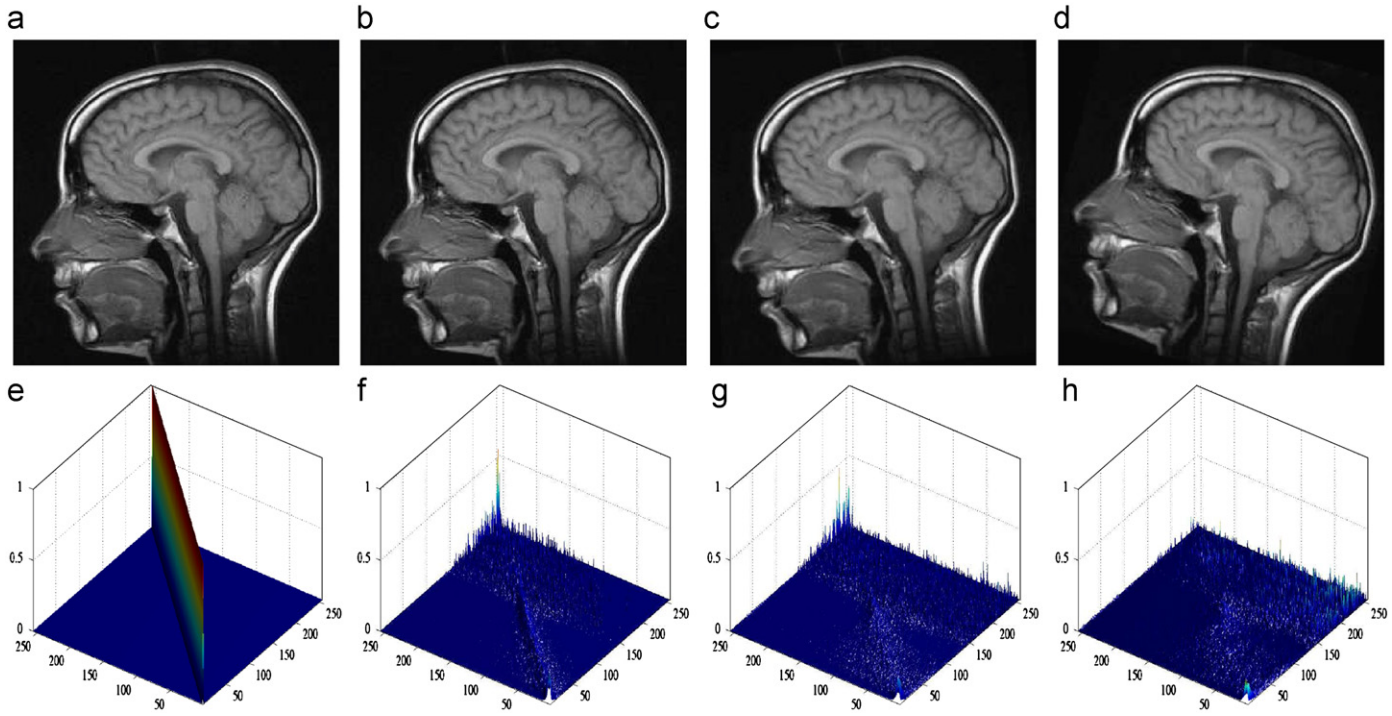


Fig. 4. 3D plots of conditional probability distributions for different transformation vectors. (a) $\ell = (0, 0, 0)$, (b) $\ell = (2, 0, 0)$, (c) $\ell = (0, 0, 5)$, and (d) $\ell = (5, 0, -15)$.

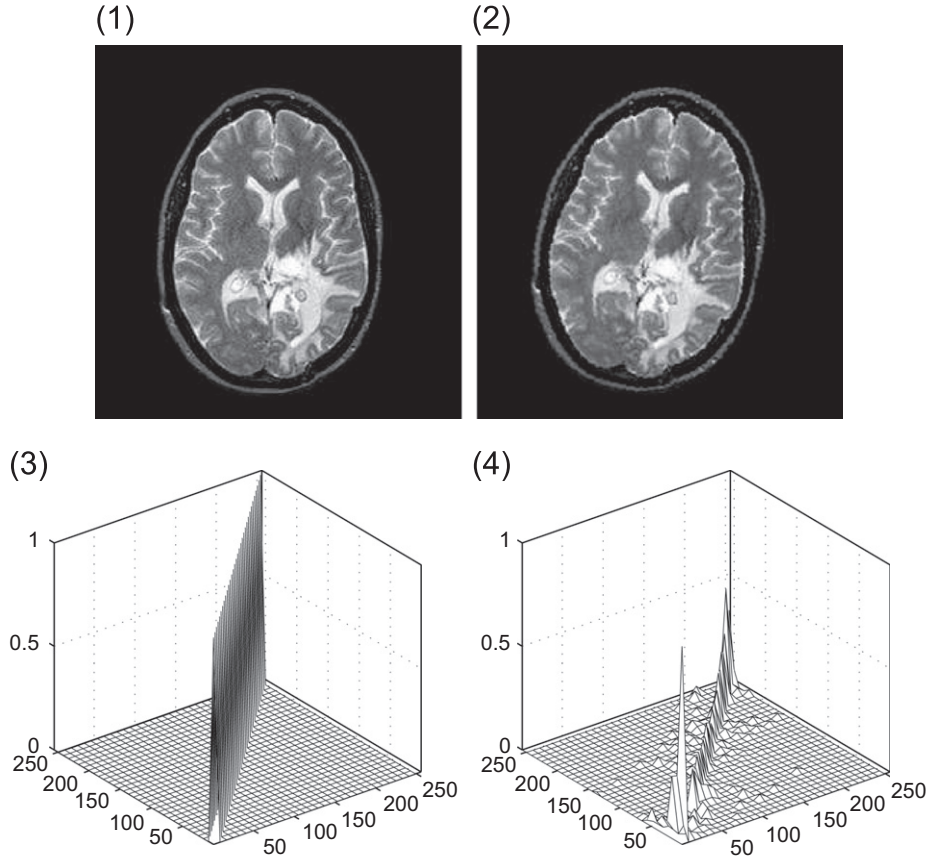


Fig. 5. Conditional probability distributions.

dispersion, information, and randomness. The maximum uncertainty or equivalently minimum information is achieved by the uniform distribution. Hence, we can think of the entropy as a measure of uniformity of a probability distribution. Consequently, when uncertainty is higher it becomes more difficult to predict the outcome of a draw from a probability distribution. A generalization of Shannon entropy is Rényi entropy [16] given by

$$R_\alpha(\boldsymbol{p}) = \frac{1}{1-\alpha}\log\sum_{j=1}^{k} p_j^\alpha, \quad \alpha \in (0,1)\cup(1,\infty). \tag{2}$$

Another important generalization of Shannon entropy is Tsallis entropy [12,17,18] given by

$$H_\alpha(\boldsymbol{p}) = \frac{1}{1-\alpha}\left(\sum_{j=1}^{k} p_j^\alpha - 1\right) = -\sum_{j=1}^{k} p_j^\alpha \log_\alpha(p_j), \tag{3}$$

where $\log_\alpha$ is the α-logarithm function defined as $\log_\alpha(x) = (1-\alpha)^{-1}(x^{1-\alpha}-1)$ for $x>0$. This generalized entropy was first introduced by Havrda and Charvát in [17], who were primarily interested in providing another measure of entropy. Tsallis, however, appears to have been principally responsible for investigating and popularizing the widespread physics applications of this entropy which is referred to nowadays as Tsallis entropy [12]. It is worth noting that for $\alpha \in (0,1]$, Rényi and Tsallis entropies are both concave functions; and for $\alpha > 1$ Tsallis entropy is also concave, but Rényi entropy is neither concave nor convex. Furthermore, both entropies tend to Shannon entropy $H(\boldsymbol{p})$ as $\alpha \to 1$, and are related by

$$H_\alpha(\boldsymbol{p}) = \frac{1}{1-\alpha}[\exp\{(1-\alpha)R_\alpha(\boldsymbol{p})\}-1].$$

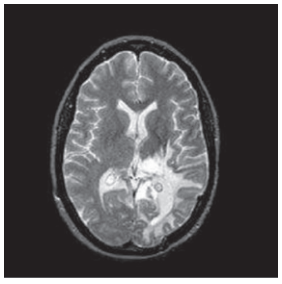
For $x,y>0$, the α-logarithm function satisfies the following property

$$\log_\alpha(xy) = \log_\alpha x + \log_\alpha y + (\alpha-1)\log_\alpha x \log_\alpha y. \tag{4}$$

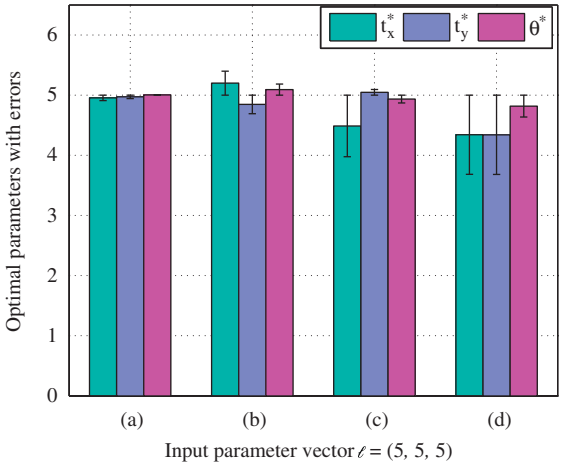
If we consider that a physical system can be decomposed in two statistical independent subsystems with probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$, then using Eq. (4) it can be shown that the joint Tsallis entropy is pseudo-additive

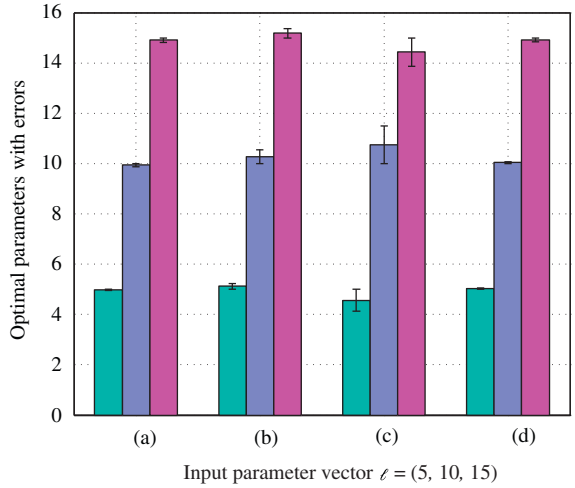
$$H_\alpha(\boldsymbol{p},\boldsymbol{q}) = H_\alpha(\boldsymbol{p}) + H_\alpha(\boldsymbol{q}) + (1-\alpha)H_\alpha(\boldsymbol{p})H_\alpha(\boldsymbol{q}),$$

whereas the joint Shannon and Rényi entropies satisfy the additivity property: $H(\boldsymbol{p},\boldsymbol{q}) = H(\boldsymbol{p}) + H(\boldsymbol{q})$, and $R_\alpha(\boldsymbol{p},\boldsymbol{q}) = R_\alpha(\boldsymbol{p}) + R_\alpha(\boldsymbol{q})$.



Reference image





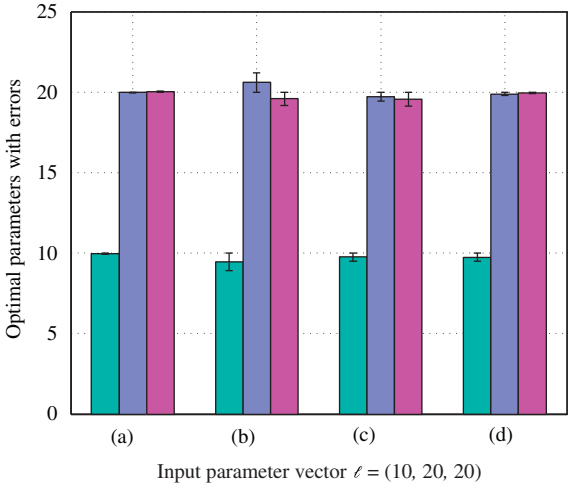


Fig. 6. Bar plots of optimal registration parameter vector $\ell^\star = (t_x^\star, t_y^\star, \theta^\star)$ with errors $\ell - \ell^\star$, using: (a) proposed method, (b) Jensen–Rényi divergence, (c) mutual information, and (d) Tsallis mutual information.

The pseudo-additivity property implies that Tsallis entropy has a nonextensive property for statistical independent systems, whereas Shannon and Rényi entropies have the extensive property (i.e. additivity). Furthermore, standard thermodynamics is extensive because of the short-range nature of the interaction between subsystems of a composite system. In other words, when a system is composed of two statistically independent subsystems, then the Boltzman–Gibbs entropy of the composite system is just the sum of entropies of the individual systems, and hence the correlations between the subsystems are not accounted for. Tsallis entropy, however, does take into account these correlations due to its pseudo-additivity property. Furthermore, many objects in nature interact through long-range interactions such as gravitational or unscreened Coulomb forces. Therefore, the property of additivity is very often violated, and consequently the use of a nonextensive entropy is more suitable for real-world applications. Fig. 2 depicts Tsallis entropy of a Bernoulli distribution $\boldsymbol{p}=(p,1-p)$, with different values of the parameter α. As illustrated in Fig. 2, the measure of uncertainty is at a minimum when Shannon entropy is used, and for $\alpha \geq 1$ it decreases as the parameter α increases. Furthermore, Tsallis entropy attains a maximum uncertainty when its exponential order α is equal to zero.

Definition 1. Let $\boldsymbol{p}_1,\boldsymbol{p}_2,\ldots,\boldsymbol{p}_n$ be n probability distributions. The Jensen–Tsallis divergence is defined as

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

where $H_\alpha(\boldsymbol{p})$ is Tsallis entropy, and $\boldsymbol{\omega}=(\omega_1,\omega_2,\ldots,\omega_n)$ is a weight vector such that $\sum_{i=1}^{n}\omega_i=1$ and $\omega_i\geq 0$.

Using the Jensen inequality, it is easy to check that the Jensen–Tsallis divergence is nonnegative for $\alpha>0$. 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 $\alpha>0$. Note that the Jensen–Shannon divergence [19] is a limiting case of the Jensen–Tsallis divergence when $\alpha\to 1$.

Unlike other entropy-based divergence measures such as the Kullback-Leibler divergence, the Jensen–Tsallis divergence has the advantage of being symmetric and generalizable to any arbitrary number of probability distributions or data sets, with a possibility of assigning weights to these distributions. Fig. 3 shows three-dimensional representations and contour plots of the Jensen–Tsallis divergence with equal weights between two Bernoulli distributions $\boldsymbol{p}=(p,1-p)$ and $\boldsymbol{q}=(q,1-q)$, for $\alpha\in(0,1)$ and also for $\alpha\in(1,\infty)$.

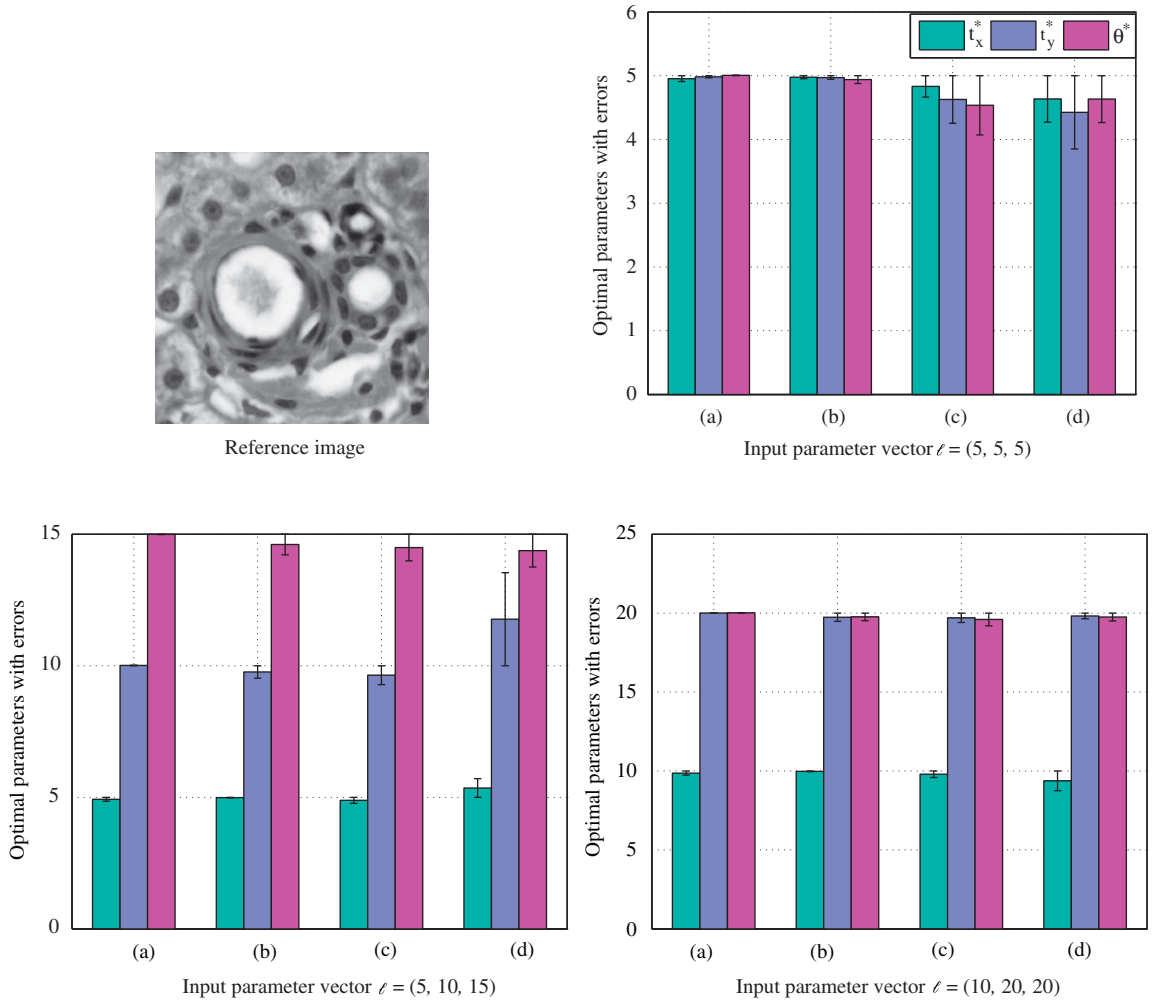


Fig. 7. Bar plots of optimal registration parameter vector $\ell^\star=(t_x^\star,t_y^\star,\theta^\star)$ with errors $\ell-\ell^\star$, using: (a) proposed method, (b) Jensen–Rényi divergence, (c) mutual information, and (d) Tsallis mutual information.

2.2. Properties of the Jensen–Tsallis divergence

The following result establishes the convexity of the Jensen–Tsallis divergence of a set of probability distributions [18].

Proposition 1. *For $\alpha \in [1,2]$, the Jensen–Tsallis divergence D_α^ω is a convex function of $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$.*

In the sequel, we will restrict $\alpha \in [1,2]$, unless specified otherwise. In addition to its convexity property, the Jensen–Tsallis divergence is an adapted measure of disparity among n probability distributions as shown in the next result.

Proposition 2. *The Jensen–Tsallis divergence D_α^ω 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. The domain of the Jensen–Tsallis divergence is a convex polytope in which the vertices are degenerate probability distributions. That is, the maximum value of the Jensen–Tsallis divergence occurs at one of the extreme points which are the degenerate distributions. □

2.3. Modified SPSA optimization algorithm

The implementation of SPSA depends on a simple "simultaneous perturbation" approximation to the gradient [15]. It uses only two measurements of the loss function in each iteration independent of the number of the problem dimension. In contrast, the standard stochastic approximation method like finite difference stochastic approximation varies the variables one at a time. If the number of terms being optimized is equal to v, then the finite-difference method takes $2v$ measurements of the objective function at each iteration.

Next we propose a modified SPSA algorithm that maximizes a real-valued loss function $\mathcal{L}(\ell)$, where ℓ denotes a v-dimensional transformation parameter vector that needs to be optimally found by maximizing $\mathcal{L}(\ell)$. The proposed SPSA algorithm starts from an initial guess of ℓ, where the iteration process depends on the above-mentioned highly efficient "simultaneous perturbation" approximation to the gradient $g(\ell) \equiv \nabla\mathcal{L}(\ell)$. It is assumed that $\mathcal{L}(\ell)$ is a differentiable function of ℓ and that the maximum point ℓ^* corresponds to a zero point of the gradient, i.e.,

$$g(\ell^*) = \nabla\mathcal{L}(\ell^*) = 0.$$

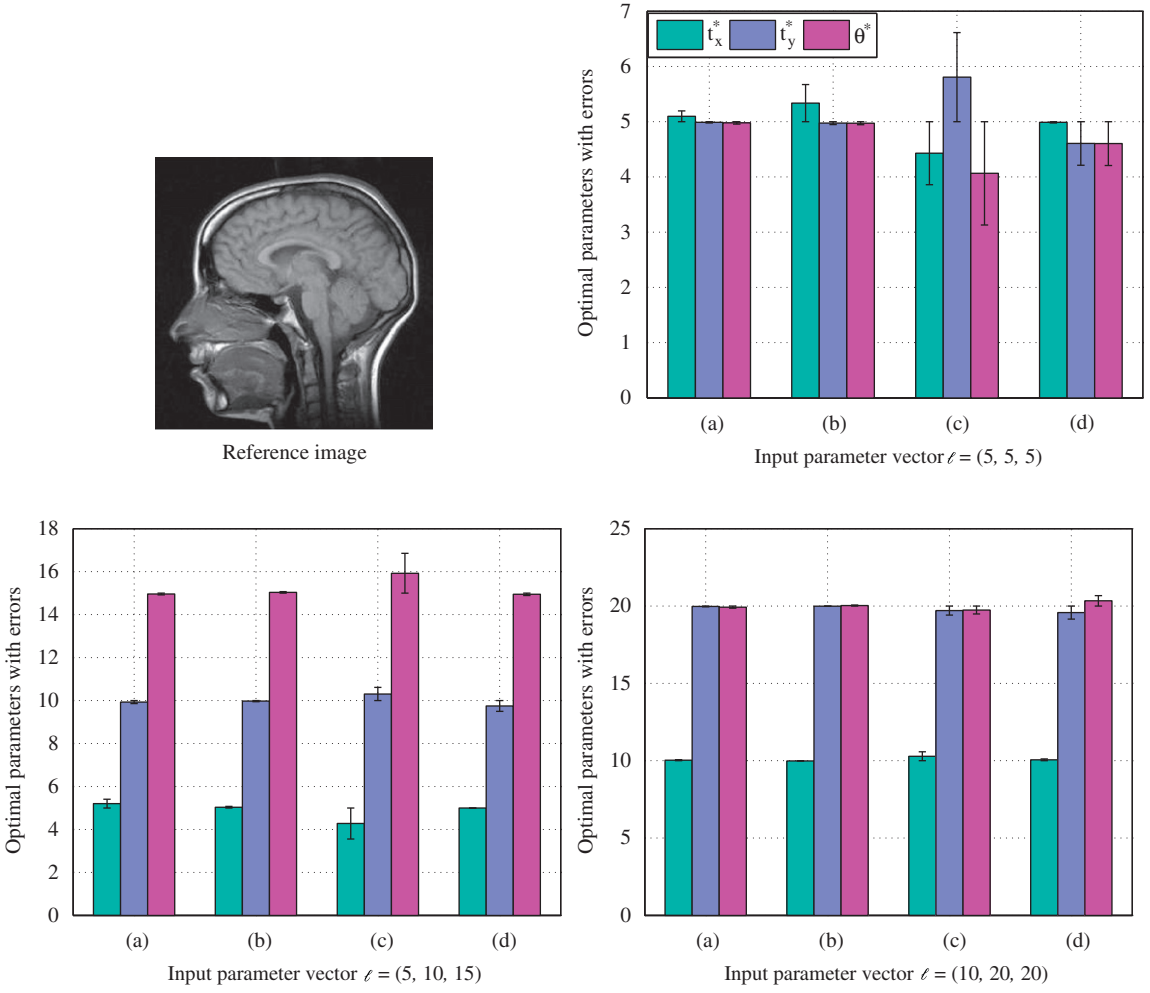


Fig. 8. Bar plots of optimal registration parameter vector $\ell^\star = (t_x^\star, t_y^\star, \theta^\star)$ with errors $\ell - \ell^\star$, using: (a) proposed method, (b) Jensen–Rényi divergence, (c) mutual information, and (d) Tsallis mutual information.

Let $y(\ell)=\mathcal{L}(\ell)+noise$, and $\hat{\ell}$ be the estimate of ℓ. Then the gradient estimate $\hat{g}(\hat{\ell})$ in the k-th iteration is given by

$$\hat{g}_k(\hat{\ell}_k)=\frac{y(\hat{\ell}_k+c_k\varepsilon_k)-y(\hat{\ell}_k-c_k\varepsilon_k)}{2c_k\varepsilon_k},$$

where c_k is the perturbation coefficient, and ε_k is the v-dimensional simultaneous perturbation vector that is Monte Carlo-generated. At the end of each iteration, the $\hat{\ell}_k$ estimate is updated using the standard stochastic approximation form

$$\hat{\ell}_{k+1}=\hat{\ell}_k+a_k\hat{g}_k(\hat{\ell}_k).$$

Note that the choice of the gain sequences a_k and c_k should satisfy some typical stochastic approximation conditions [15].

3. Proposed method

Our proposed approach may now be described as follows: Given two images that need to be registered, we first compute their conditional intensity probabilities and the Jensen–Tsallis divergence between them. Then we optimize this divergence measure using the modified SPSA algorithm.

Without loss of generality, we consider a Euclidean transformation Φ_ℓ with a parameter vector $\ell=(\boldsymbol{t},\theta)$, i.e. a transformation with translation parameter vector $\boldsymbol{t}=(t_x,t_y)$, and a rotation parameter θ. In other words, for an image pixel location $\boldsymbol{x}=(x,y)$ the Euclidean transformation is defined as $\Phi_\ell(\boldsymbol{x})=R\boldsymbol{x}+\boldsymbol{t}$, where R is a rotation matrix given by

$$R=\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}.$$

Denote by $\mathcal{X}=\{x_1,x_2,\ldots,x_n\}$ and $\mathcal{Y}=\{y_1,y_2,\ldots,y_n\}$ the sets of pixel intensity values of the reference image $I_1(\boldsymbol{x})$ and the transformed target image $I_2(\Phi_\ell(\boldsymbol{x}))$, respectively. Let X and Y be two random variables taking values in $\mathcal{X}$ and $\mathcal{Y}$.

The proposed approach consists of the following main steps:

(i) Find the conditional intensity probabilities

$$\boldsymbol{p}_i=\boldsymbol{p}_i(I_2(\Phi_\ell(\boldsymbol{x}))|I_1(\boldsymbol{x}))=(p_{ij})_{j=1,\ldots,n}\quad \forall i=1,\ldots,n,$$

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

(ii) Find the optimal parameter vector $\ell^\star=(\boldsymbol{t}^\star,\theta^\star)$ of the Jensen–Tsallis objective function

$$\ell^\star=\arg\max_\ell D_\alpha^\omega(\boldsymbol{p}_1,\ldots,\boldsymbol{p}_n) \tag{5}$$

using the modified SPSA optimization algorithm (Fig. 4).

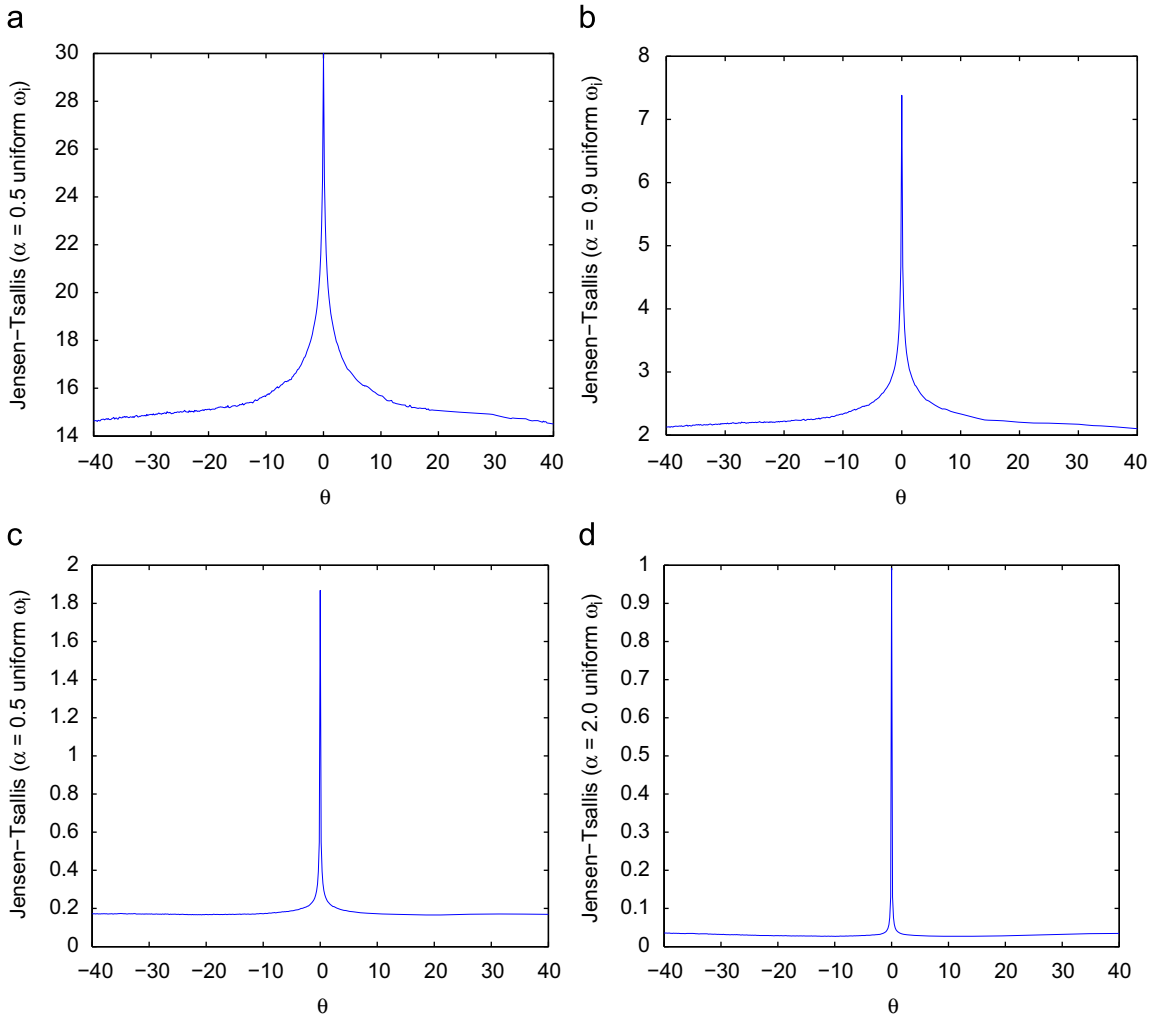


Fig. 9. Jensen–Tsallis divergence with uniform weight.

Note that if the images I_1 and I_2 are exactly matched, then $\boldsymbol{p}_i = (\delta_{ij})$ and by Proposition 2, the Jensen–Tsallis divergence is therefore maximized. Fig. 5(1) and (2) shows two MRI images in which the misalignment corresponds to a clockwise rotation with an angle $\theta = 10^{\circ}$. The conditional probability distributions $\{\boldsymbol{p}_i\}$ are crisp, as shown in Fig. 5(3), when the two images are aligned, and dispersed, as depicted in Fig. 5(4), when they are not matched.

Also, it is worth pointing out that if $\alpha = 1$ and $\omega_i = P(X = x_i)$ then the Jensen–Tsallis divergence becomes mutual information, indicating that the Jensen–Tsallis divergence induces a dissimilarity measure that provides a more general framework for the image registration problem.

4. Experimental results

We tested the performance of the proposed entropic image registration method on a variety of images. In all experiments we used an entropic index $\alpha = 2$ and the normalized histogram as the weight vector $\boldsymbol{\omega}$ for the Jensen–Tsallis divergence. In our first experiment, we applied a Euclidean transformation Φ_ℓ with different values of the parameter vector $\ell = (t_x, t_y, \theta)$ to the three reference medical images shown in Figs. 6–8. And, we used the modified SPSA algorithm to find the optimal parameter vector $\ell^* = (t_x^\star, t_y^\star, \theta^\star)$. We also compared the image alignment results of the proposed approach to existing image registration techniques based on the mutual information [3], Tsallis mutual information [13], and Jensen–Rényi divergence [7]. The output registration results are shown in Figs. 6–8, where the absolute differences $|t_x - t_x^\star|$, $|t_y - t_y^\star|$, and $|\theta - \theta^\star|$ between the true and the estimated transformation parameters are also displayed as error bars for three different transformation parameter vectors $\ell = (5,5,5)$, $\ell = (5,10,15)$, and $\ell = (10,20,20)$. From these figures, it is clear that the estimated values of the transformation parameters indicate the effectiveness and the registration accuracy of the proposed algorithm. Amongst the other methods, we noticed that the Tsallis mutual information approach performs relatively well at higher values of the rotation angle, but poorly at higher values of the translation parameters compared to the proposed approach. Moreover, the much better performance of our method is in fact consistent with a variety of images used for experimentation.

In the next experiments, we examine the effects of the values of the weight vector $\boldsymbol{\omega}$ and the entropic index α

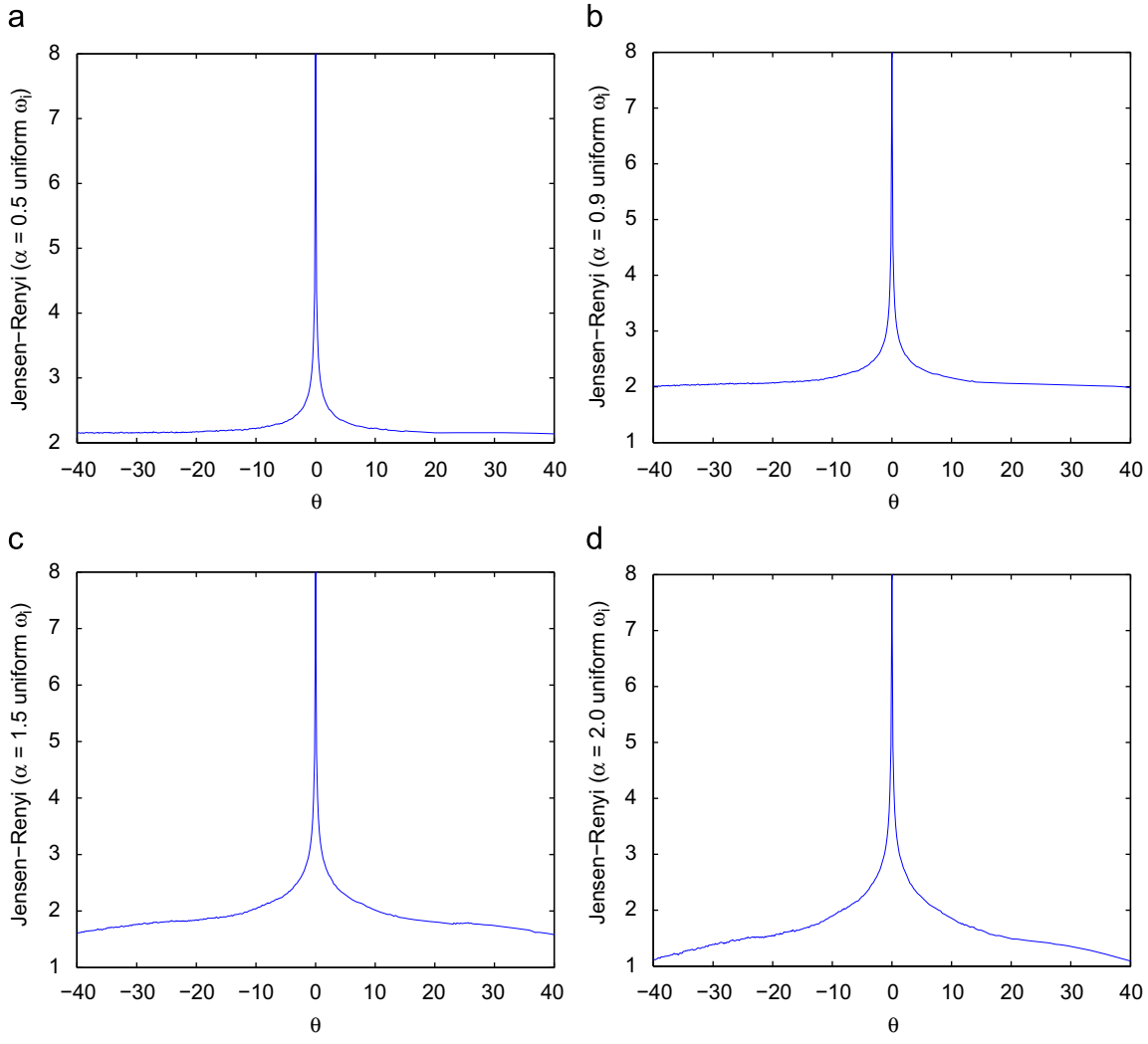


Fig. 10. Jensen–Rényi divergence with uniform weight.

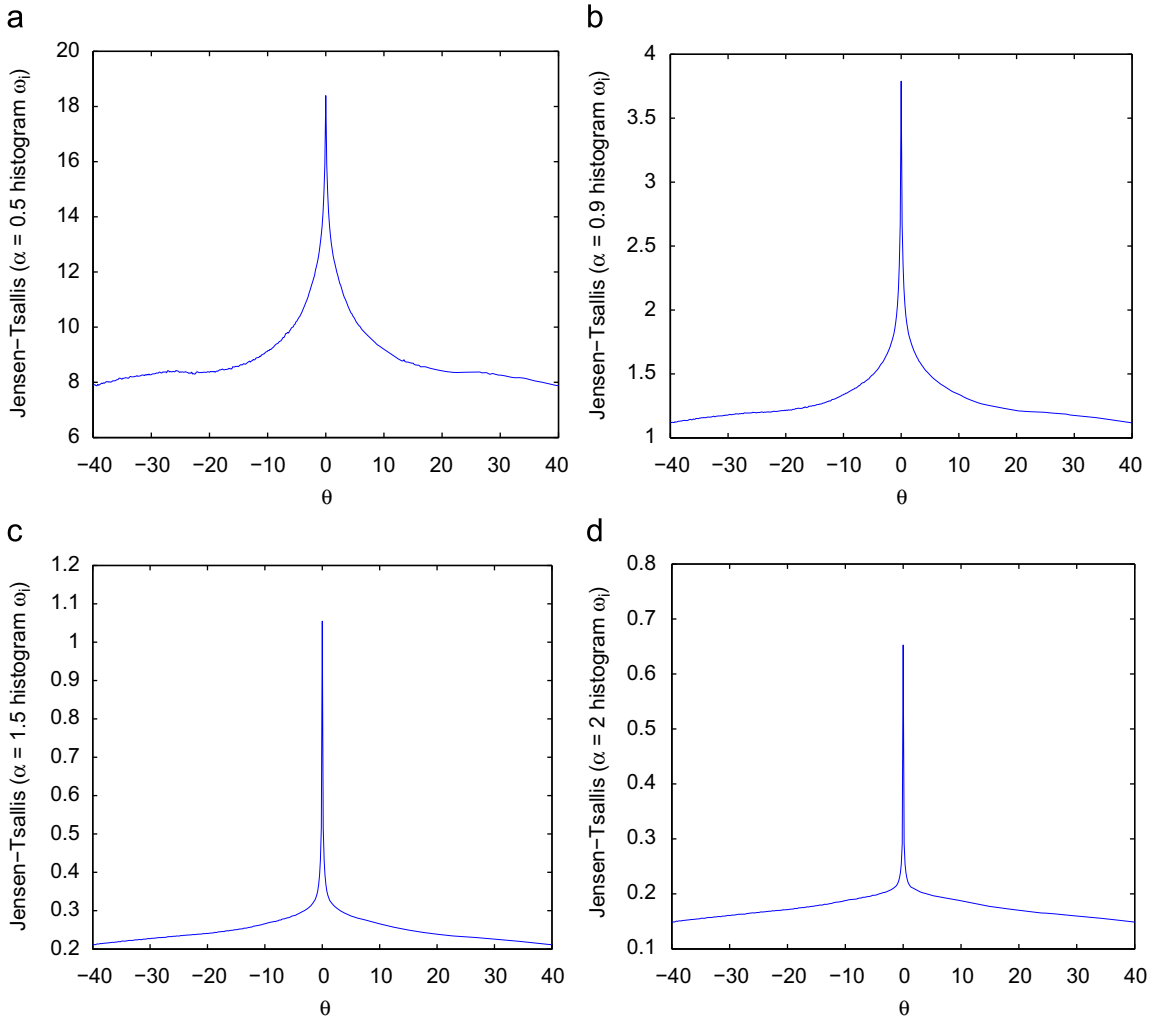


Fig. 11. Jensen–Tsallis divergence with normalized histogram weight.

on the performance of the proposed image registration approach.

4.1. Effect of weight on the Jensen–Tsallis divergence

Figs. 9 and 10 show the plots for the divergence values of Jensen–Tsallis and Jensen–Rényi, respectively, in the case of a uniform weight $\omega_i = 1/n$. As shown in Fig. 10, the Jensen–Rényi divergence has the same maximum value given different values of α, whereas the maximum value of the Jensen–Tsallis divergence drops significantly when the value of α increases. Also, it is worth noting that the output of the Jensen–Tsallis divergence shows a sharp impulse located at where the images are aligned, and a uniform value anywhere else if a uniform weight $\omega_i = 1/n$ is used and also if α is larger than 1. This property indicates a dramatic change of the gradient. Indeed, through extensive experiments we noticed that the modified SPSA algorithm experiences difficulty in converging to the optimal solution when the initial guess of ℓ is not in a small range where a noticeable gradient change is observed. To circumvent this problem, we used the normalized histogram of the reference image I_1 as the weight vector $\boldsymbol{\omega}$ instead of a uniform weight. The Jensen–Tsallis and Jensen–Rényi divergences with the normalized histogram weights are shown in Figs. 11 and 12, respectively.

4.2. Effect of α on the Jensen–Tsallis divergence

Most medical imaging methods, for instance, produce a full three-dimensional (3D) volume, and the medical scans are viewed as a series of superposed two-dimensional (2D) slices of these full 3D volume. The MRI 3D volume of a healthy patient, shown in Fig. 13(a), consists of 27 horizontal slices and each slice is 128×128 pixels. To examine the effect of the entropic index α, we applied the proposed approach to two horizontal slices of this MRI 3D volume: the reference image and the misaligned image, which are shown on the left and right-hand sides of Fig. 13(b), respectively. Figs. 14 and 15 display the output results of Jensen–Rényi and Jensen–Tsallis divergences with uniform weight and also with normalized histogram weight.

5. Conclusions

We proposed an entropic image alignment method by optimizing a generalized divergence measure using a modified simultaneous perturbation stochastic approximation algorithm. The registration is achieved by finding the optimal Euclidean transformation parameters that maximize the Jensen–Tsallis divergence. The main advantages of the proposed approach are: (i) Jensen–Tsallis divergence is symmetric,

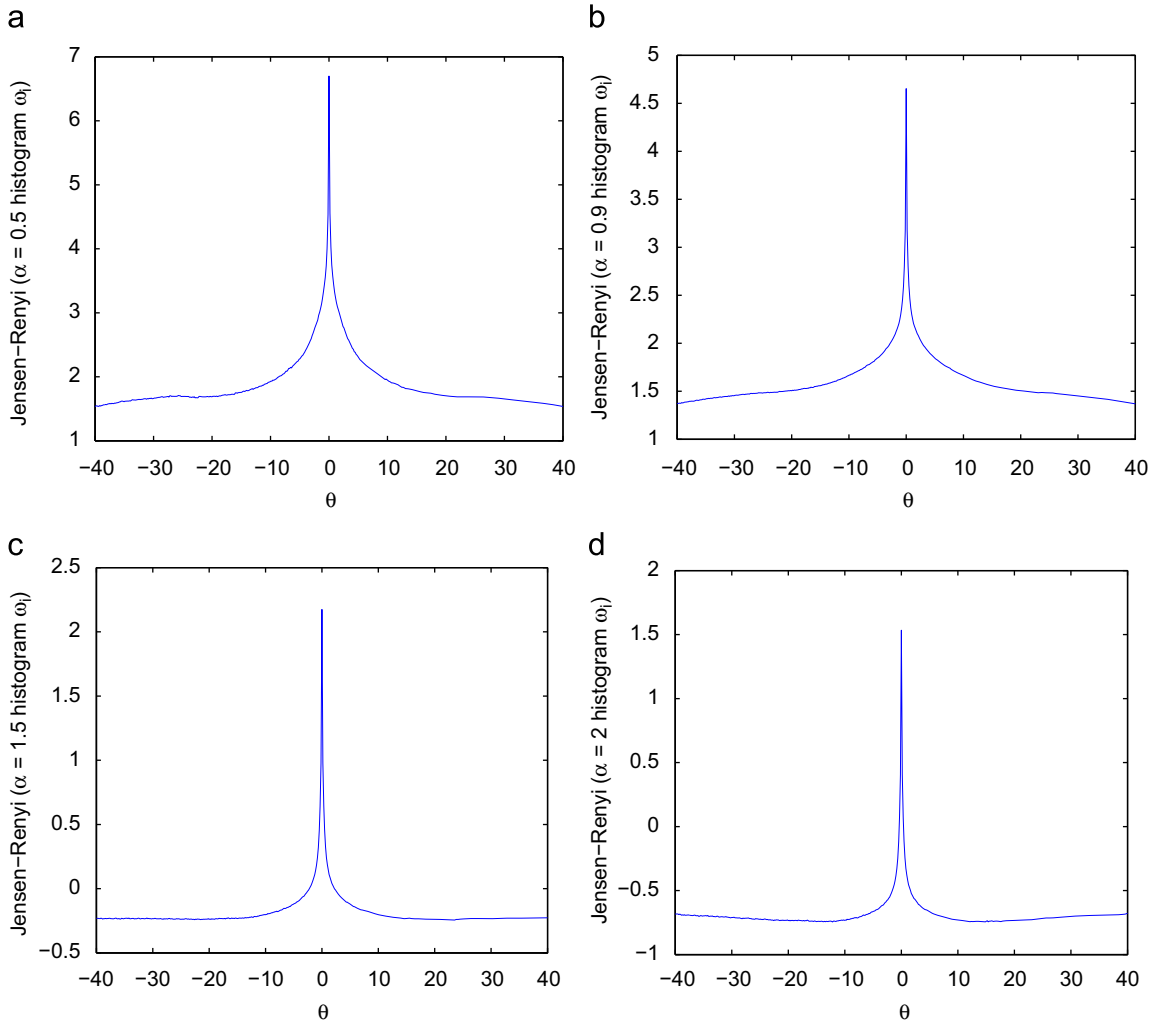


Fig. 12. Jensen–Rényi divergence with normalized histogram weight.

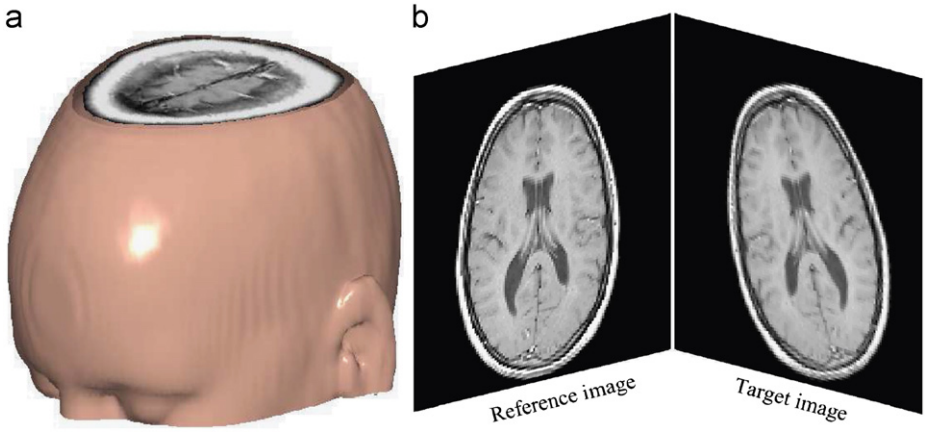


Fig. 13. (a) MRI 3D volume and (b) reference and target images for testing the effect of the entropic index.

convex, theoretically upper-bounded, and quantifies efficiently the statistical dissimilarity between the reference image and the transformed target image, and (ii) the experimental results provide accurate registration results in comparison with existing techniques. Our future goal is to incorporate the prior information on the joint intensity histogram between the images being registered for a more robust image alignment.

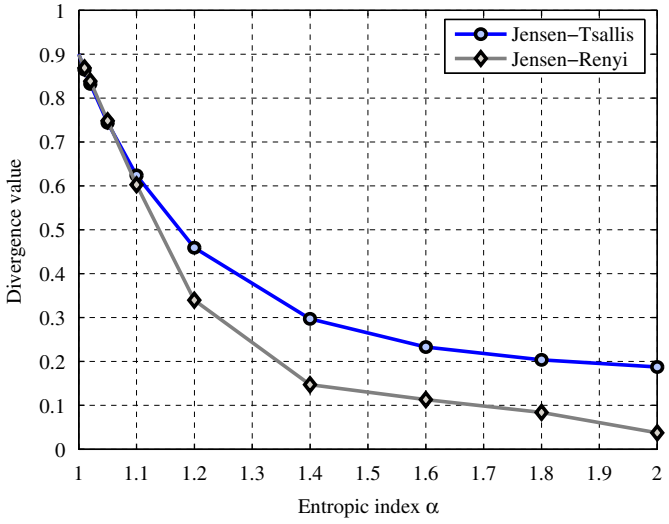


Fig. 14. Divergence values with uniform weight ω_i.

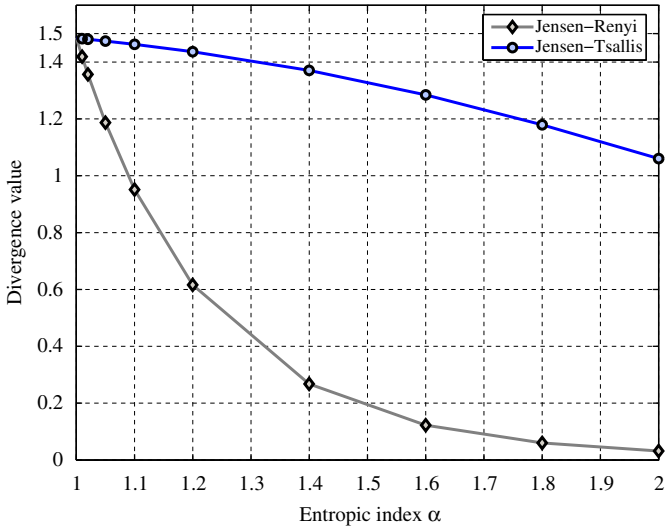


Fig. 15. Divergence values with normalized histogram weight ω_i.

References

[1] Hajnal J, Hill D, Hawkes D, editors. Medical image registration. CRC Press LLC; 2001.

[2] Goshtasby AA. 2-D and 3-D image registration for medical, remote sensing, and industrial applications. Wiley Publishers; 2005.

[3] Viola P, Wells WM. Alignment by maximization of mutual information. International Journal of Computer Vision 1997;24(2):154–73.

[4] Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multimodality image registration by maximization of mutual information. IEEE Transactions on Medical Imaging 1997;16(2):187–98.

[5] Hero AO, Ma B, Michel O, Gorman J. Applications of entropic spanning graphs. IEEE Signal Processing Magazine 2002;19(5):85–95.

[6] Pluim JPW, Maintz JBA, Viergever MA. f-Information measures in medical image registration. IEEE Transactions on Medical Imaging 2004;23(12): 1508–16.

[7] He Y, Ben Hamza A, Krim H. A generalized divergence measure for robust image registration. IEEE Transactions on Signal Processing 2003;51(5): 1211–20.

[8] Corici D, Astola J. Segmentation of DNA into coding and noncoding regions based on recursive entropic segmentation and stop-codon statistics. EURASIP Journal on Advances in Signal Processing 2004;1:81–91.

[9] Hibbard LS. Region segmentation using information divergence measures. Medical Image Analysis 2004;8(3):233–44.

[10] Chiang M-C, Dutton RA, Hayashi KM, Lopez OL, Aizenstein HJ, Toga AW, et al. 3D pattern of brain atrophy in HIV/AIDS visualized using tensor-based morphometry. NeuroImage 2007;34(1):44–60.

[11] Karakos D, Khudanpur S, Eisner J, Priebe CE. Iterative denoising using Jensen–Rényi divergences with an application to unsupervised document categorization. In: Proceedings of the IEEE international conference on acoustics, speech and signal processing, Honolulu, Hawaii, 2007.

[12] Tsallis C. Possible generalization of Boltzmann–Gibbs statistics. Journal of Statistical Physics 1988;52:479–87.

[13] Martin S, Morison G, Nailon W, Durrani T. Fast and accurate image registration using Tsallis entropy and simultaneous perturbation stochastic approximation. Electronic Letters 2004;40(10):595–7.

[14] Aguilera M, Ben Hamza A. An information-theoretic approach to georegistration of digital elevation maps. In: Proceedings of the IEEE conference on computer and robot vision, Quebec City, Canada, 2006.

[15] Spall JC. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. IEEE Transactions on Automatic Control 1992;37(3):332–41.

[16] Rényi A. On measures of entropy and information. Selected Papers of Alfréd Rényi 1961;2:525–80.

[17] Havrda ME, Charvát F. Quantification method of classification processes: concept of structural α–entropy. Kybernitica 1967;3:30–5.

[18] Burbea J, Rao CR. On the convexity of some divergence measures based on entropy functions. IEEE Transactions on Information Theory 1982;28(3): 489–95.

[19] Lin J. Divergence measures based on the Shannon entropy. IEEE Transactions on Information Theory 1991;37(1):145–51.