Multimodal registration of remotely sensed images based on Jeffrey's divergence

CrossMark

Xiaocong Xu [a], Xia Li [a,*], Xiaoping Liu [a,*], Huanfeng Shen [b], Qian Shi [a]

[a] School of Geography and Planning, and Guangdong Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou 510275, PR China
[b] School of Resource and Environmental Science, Wuhan University, Wuhan 430079, PR China

A B S T R A C T

Entropy-based measures (e.g., mutual information, also known as Kullback-Leiber divergence), which quantify the similarity between two signals, are widely used as similarity measures for image registration. Although they are proven superior to many classical statistical measures, entropy-based measures, such as mutual information, may fail to yield the optimum registration if the multimodal image pair has insufficient scene overlap region. To overcome this challenge, we proposed using the symmetric form of Kullback-Leiber divergence, namely Jeffrey's divergence, as the similarity measure in practical multimodal image registration tasks. Mathematical analysis was performed to investigate the causes accounting for the limitation of mutual information when dealing with insufficient scene overlap image pairs. Experimental registrations of SPOT image, Landsat TM image, ALOS PalSAR image, and DEM data were carried out to compare the performance of Jeffrey's divergence and mutual information. Results indicate that Jeffrey's divergence is capable of providing larger feasible search space, which is favorable for exploring optimum transformation parameters in a larger range. This superiority of Jeffrey's divergence was further confirmed by a series of paradigms. Thus, the proposed model is more applicable for registering image pairs that are greatly misaligned or have an insufficient scene overlap region.

© 2016 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

1. Introduction

Image registration is an important problem and a fundamental task in remotely sensed image analysis. Remotely sensed images collected by disparate sensors have different radiometric properties such as physical magnitudes, spectral bands, and image coherency (microwave and optical images) (Inglada and Giros, 2004). These properties provide various pieces of complementary information on terrestrial objects that we are interested in. By integrating multiple images, information obtained from separated images can be combined into one for further processing and analysis such as image fusion (Jiang et al., 2012; Luo et al., 2013; Zhang et al., 2012), multisource data assimilation (Bach and Mauser, 2003; Launay and Guerif, 2005), and land cover change detection (Almeida-Filho et al., 2005; Liu et al., 2010, 2014; Servello et al., 2010). All these analyses require accurate multimodal image registration in advance.

Image registration usually refers to the process of bringing several images into spatial alignment that involve the same terrestrial target but are collected under different viewing conditions and/or by using separate imaging devices (Kern and Pattichis, 2007). In general, an image registration model is composed of three parts: a spatial transformation, a similarity measure, and an optimization procedure (Ardizzone et al., 2009). If a registration model is applied to an image pair, we usually designate one image as the reference and apply a spatial transformation to the other image (floating image) to bring them into alignment. Throughout the registration, a similarity measure (e.g., mutual information (Collignon et al., 1995), normalized cross correlation (Lemieux et al., 1994), etc.) is defined to verify the quality of alignment between the image pair. Then, an optimization algorithm (gradient descent algorithm (Klein et al., 2009), genetic algorithm (Silva et al., 2005), etc.) is carried out to search for the extreme of the similarity measure in the transformation space, which indicates the perfect alignment of the image pairs.

During the registration process, choosing the appropriate similarity measure is crucial because it significantly affects the accuracy and robustness of the registration result. In an ideal image registration model, the similarity measure is supposed to obtain

* Corresponding authors.
  E-mail addresses: xuxiaoc@mail3.sysu.edu.cn (X. Xu), lixia@mail.sysu.edu.cn (X. Li), liuxp3@mail.sysu.edu.cn (X. Liu), shenhf@whu.edu.cn (H. Shen), shiqian@whu.edu.cn (Q. Shi).

an extreme value *if and only if* the two images are perfectly aligned. Currently, numerous studies on entropy-based similarity measures have been conducted in the field of automatic image registration (Chiang et al., 2006; He et al., 2003; Martin and Durrani, 2007; Pluim et al., 2004; Wachowiak et al., 2003a, 2003b). In fact, most of these entropy-based measures are derived from f-divergences (Ali and Silvey, 1966; Pardo and Vajda, 2003; Pluim et al., 2004; Topsoe, 2000). Statistically, f-divergence family members arise naturally as distance notions (or error exponents) between two distributions. Mutual information (Collignon et al., 1995; Viola and Wells, 1997), referred to as KL-divergence in probability theory, is one of them. It is proven that mutual information is superior to many classical statistical measures for its minimal assumptions on the correspondence between image intensities as well as its fully automated implementation (Seppa, 2008). As long as a certain co-dependence exists between an image pair, mutual information can obtain a peak value when the two images are registered. Since then, mutual information has drawn tremendous interest and has been adopted for numerous applications (Pluim et al., 2000; Studholme et al., 1995, 1997). In addition, some modified versions of mutual information, such as normalized mutual information (Studholme et al., 1999) and weight mutual information (Li et al., 2008), have been developed to improve the performance of multimodal remotely sensed image registration and achieve outstanding results (Chen and Varshnev, 2004; Chen et al., 2003a, 2003b; Cole-Rhodes et al., 2003; Inglada and Giros, 2004; Inglada et al., 2007; Johnson et al., 2001; Xie et al., 2003).

These abovementioned models mainly focus on pursuing sub-pixel registration accuracy by designing better similarity measures (He et al., 2003; Liang et al., 2014; Karantzalos et al., 2014; Martin and Durrani, 2007; Pluim et al., 2004) or applying advanced interpolation methods (e.g., PVE (Hasan et al., 2012; Maes et al., 1997), GPVE (Chen and Varshney, 2003), and HPV (Lu et al., 2008)). Though such models can register image pairs accurately, they are only applicable to image pairs with sufficient scene overlap region (the common terrestrial region that captured by the multiple devices; see Section 3.3 for details). In practical multimodal registration tasks, the floating images and reference one are usually captured by various sensors. The scene overlap region between these images is usually insufficient. During the parameter searching process, the overlap region[1] between the image pairs keeps changing when trying different transformation parameters, and equals to the scene overlap region only if the two images are perfectly aligned. In order to find out the optimum transformation parameter, the search space has to cover a large range to ensure the optimum transformation parameter is within it. Inevitably, there exist some transformation parameters in the search space that result in even smaller overlap regions between the images than the scene overlap regions. The smaller overlap region between the image pairs, the more unstable the mutual information. In such cases, the maximum value of mutual information does not correspond to the optimum registration, but rather a certain transformation parameter that leads to small overlap between the images. Moreover, images captured via different sensors are different in coherency, which introduces random interference during the estimation of similarity measures and consequently exacerbate the uncertainty of the similarity measure. These challenges must be reckoned with when applying an automatic registration model to solve *practical multimodal registration* tasks. Otherwise, human intervention in model

initialization and search space restriction are needed to ensure that the optimization algorithm only searches around the optimal transformation parameters

Our paper aims to solve the abovementioned challenges. We analyzed the limitations of mutual information for remotely sensed image registration, particularly in these cases where multimodal image pairs were involved with an insufficient scene overlap region. Mathematical analysis and experiments were then carried out to investigate the underlying causes accounting for these limitations. On this basis, the symmetric form of mutual information, namely Jeffrey's divergence, was adopted as the similarity measure to overcome the drawbacks of mutual information in multimodal image registration application. By providing larger feasible search space, registration model based on Jeffrey's divergence is more capable of registering multimodal image pairs that are greatly misaligned or with insufficient scene overlap. Mutual information will be used as the basis upon which to test the performance of the proposed registrations models since it is a well-accepted similarity measure from which numerous entropy-based similarity measures have been derived.

2. Image registration based on intensity

Existing image registration models consist of three categories concerning different information they utilize: feature, intensity, and transform domains (Lin et al., 2009). Feature-based models are data-dependent in that different image data may contain features in different aspects. Extracting the same features in multimodal images and matching them together is usually difficult. Models based on transform domains are robust in translation misalignment, but are undesirable for dealing with rotation and scaling. Compared with registration models based on features and transform domains, intensity-based registration models are easy to carry out and are free of error propagation (Zitova and Flusser, 2003), even though sometimes they are quite time-consuming. They simply utilize the intensity information lying in the whole image to build a statistical relation and describe the similarity of the image pair, which makes it applicable to multimodal image registration.

2.1. Mathematical description of intensity-based registration

To clarify the mathematical description of the image registration model, we have the following definitions according to previous researches: (1) we define the image histogram of the overlap region covered by the aligned image pair as a random variable $\mathbf{I}$, and use $\mathbf{I}_R$ and $\mathbf{I}_F$ to denote the random variables of reference image and floating image respectively; (2) the sample spaces for $\mathbf{I}_R$ and $\mathbf{I}_F$ are *theoretically* defined as image pixels covered by the overlap region of the aligned image pair (corresponding to the scene overlap region). However, the overlap region of the aligned image pair is actually unknown in the image registration process (this is what the registration model attempts to determine). Therefore, in this paper, the sample space is defined as image pixels covered by the overlap region of the reference image and the transformed floating image. The extra space around the images with 0-value is introduced to keep the sample populations proximately consistent.

In general, a typical image registration methodology is equivalent to a function optimization problem (Modersitzki, 2004). The optimization seeks the maximum similarity of an image pair by determining a mathematical mapping $\boldsymbol{S} : \boldsymbol{\mu} \rightarrow \boldsymbol{s}$ from the spatial transformation parameter space $\boldsymbol{\mu}$ to the similarity measure space $\boldsymbol{s}$. So the registration model can be mathematically expressed as

---

[1] Note that "overlapping region of the image pairs" is different from "scene overlapping region". Specifically, "overlapping region of two images" refers to overlap of the floating image and the reference one. It keeps changing during the parameters searching process in registration. However, "scene overlapping region" is the common terrestrial region that captured by the multiple devices. It corresponds to the "overlapping region of two images" only if these two images are perfectly aligned.

$$T^* = \arg\max_{\mu}\{S_{\mu}[\mathbf{I}_R(\boldsymbol{X}), \mathbf{I}_F(\boldsymbol{T}_{\mu} \circ \boldsymbol{X})]\} \tag{1}$$

where $\mathbf{I}_R(\boldsymbol{X})$ and $\mathbf{I}_F(\boldsymbol{X})$ are intensity distributions that denote the reference image and the floating image respectively, which vary with their spatial coordination $\boldsymbol{X}$; $\boldsymbol{T}_{\mu}$ is a spatial transformation model characterized by a parameter vector $\boldsymbol{\mu}$, which will be applied to the coordinate of every pixel in $\mathbf{I}_F(\boldsymbol{X})$ to map the floating image into the transformed image $\mathbf{I}_F(\boldsymbol{T}_{\mu} \circ \boldsymbol{X})$; $\boldsymbol{S}_{\mu}(\cdot)$ is the objective function (similarity) that indicates the mapping from the transformation parameter vector $\boldsymbol{\mu}$ to the similarity measure $\boldsymbol{s}$; finally, $\boldsymbol{T}^*$ is the optimum transformation that generates the maximum similarity, which (theoretically) aligns the floating image to the reference image perfectly.

2.2. Image statistical model and registration implementation

According to the intensity-based registration model depicted in Section 2.1, the image distribution $\mathbf{I}(\boldsymbol{X})$ must be estimated before we can obtain the similarity measure. In previous studies, an image histogram, which plots the pixel number of each intensity level, was adopted to denote the statistical characteristics of remotely sensed images (Richards and Jia, 1999). With regard to a specific terrestrial region, histograms estimated from multimodal images share some common characteristics (e.g., having several peaks emerges at different gray values). These common characteristics imply that a certain kind of co-dependence exist among these images. In other words, intensity values populating at the same geographical position can be linked with one another via a specific spatial mapping.

However, an image histogram cannot be directly applied to image registration because it does not contain any spatial relationship between the image pair within the same coordinate system. Moreover, image registration is only concerned with the similarity of the scene overlap region, whereas an image histogram accounts for the statistical characteristics of the whole image. Several works (Gao et al., 2008; Inglada and Giros, 2004; Pluim et al., 2000; Thévenaz and Unser, 2000) have adopted the joint histogram of the image pair to embed spatial information into an intensity distribution. In a joint histogram, two one-dimensional histograms are combined into one by merging the coordinates of both images. Consequently, the joint histogram of an image pair is a bi-dimensional histogram, where two axes represent the intensity level of two images and its elements indicate the distribution of every intensity pair that shares the same coordinate (Fig. 1).

Mathematically, the joint histogram of an image pair can be presented as a two-dimensional matrix:

$$\mathbf{JH} = [c_{ij}] \quad i \leqslant M \quad j \leqslant N \tag{2}$$

where M and N are the total intensity levels of two images respectively. Each value c_{ij} located at coordinate (i,j) indicates that there are c_{ij} corresponding intensity pairs across the entire overlap region of two images, having intensity value i in the first image and intensity value j in the second image. Several examples of joint histograms are illustrated in Fig. 1. Fig. 1-d is a special case, estimated from two images that are identical and perfectly aligned. In this joint histogram, all intensity pairs are distributed along the diagonal because each intensity value in one image is equal to the corresponding value in the other image if these two images are perfectly aligned. There exists a one-to-one mapping that strictly maps each pixel value in one image to the corresponding pixel value in another image. Fig. 1-e illustrates the joint histogram of different spectral bands that are imaged at the same time by one sensor (SPOT Band4 and Band1). Fig. 1-f is the joint histogram of two images that are captured by different sensors (SPOT Band4 and Pal-SAR). If the multi-spectral image pair is aligned together, the joint

histogram shows a significant banded structure along the diagonal (Fig. 1-e). Instead of a strict one-to-one mapping, a many-to-many relationship exists between the intensities of two images. As to multimodal image registration, there will be no direct relationship between the intensities of the two images. Intensity pairs will no longer aggregate on the diagonal, but will concentrate on certain regions even if the image pair is perfectly aligned (Fig. 1-f).

Generally, intensity pairs tend to be more dispersive if a greater shift occurs between the image pair. In an image, polygon-shaped terrestrial objects take up the majority of the pixels. A greater shift will reduce the overlap area of the same terrestrial patches, which results in the diversification of the intensity pair and leads to a more dispersive distribution in the joint histogram. Thus, by measuring the distribution of intensity pairs in the joint histogram (dispersion or aggregation), we can deduce the quantity of alignment and determine the optimum transformation parameter set that leads to a perfect alignment of the floating and reference images.

Excluding some extreme cases, most intensity-based multimodal image registration methods are based on the assumption that perfect alignment corresponds to the most aggregated distribution intensity pairs in the joint histogram. Hence, the similarity measure, which essentially measures the distance between distributions of intensity pairs, indirectly reflects the registration result. Robust multimodal image registration requires an appropriate similarity measure capable of withstanding external interference and indicating the quality of alignment between image pairs.

3. Mathematical analysis of mutual information's limitation

According to Section 2, a similarity measure determines the accuracy and robustness of registration models. Mutual information, as a similarity measure, has been widely used in intensity-based multimodal image registration. In this section, we will present the basic concept of mutual information and its origin, f-divergence. The problems of mutual information in intensity-based image registration will then be analyzed, including its sensitivities to the image overlap ratio.

3.1. Definition of f-divergence

In probability theory, f-divergence is a function $D_f(\boldsymbol{P}\|\boldsymbol{Q})$ that establishes a "distance" between one probability distribution $\boldsymbol{P}$ and another probability distribution $\boldsymbol{Q}$ on a statistical manifold (Ali and Silvey, 1966). In other words, f-divergence indicates the discrimination between two distributions. Since the general notion of f-divergence was first introduced and thoroughly investigated by Csiszár (1963) and Ali and Silvey (1966), it is also known as Csiszár f-divergences or Ali-Silvey distances.

Specifically, for two arbitrary distributions $\boldsymbol{P}$ and $\boldsymbol{Q}$ defined over a space $\boldsymbol{\Omega}$, where $\boldsymbol{P}$ is absolutely continuous with respect to $\boldsymbol{Q}$, f-divergence from $\boldsymbol{P}$ to $\boldsymbol{Q}$ can be defined as (Ali and Silvey, 1966)

$$D_f(\boldsymbol{P}\|\boldsymbol{Q}) = E_{\boldsymbol{Q}}\left[f\left(\frac{d\boldsymbol{P}}{d\boldsymbol{Q}}\right)\right] = \int_{\boldsymbol{\Omega}} f\left(\frac{d\boldsymbol{P}}{d\boldsymbol{Q}}\right)d\boldsymbol{Q} \tag{3}$$

where $f(\cdot)$ is a continuous convex function that satisfies (1) $f : (0, +\infty) \rightarrow (-\infty, +\infty]$, (2) finite on $(0, +\infty)$, (3) strictly convex at some points $x \in (0, +\infty)$; and (4) $f(1) = 0$.

If $\boldsymbol{P}$ and $\boldsymbol{Q}$ are both absolutely continuous with respect to a reference distribution $\boldsymbol{\mu}$ on space $\boldsymbol{\Omega}$, such that $d\boldsymbol{P} = p d\boldsymbol{\mu}$ and $d\boldsymbol{Q} = q d\boldsymbol{\mu}$, then

$$D_f(\boldsymbol{P}\|\boldsymbol{Q}) = \int_{\boldsymbol{\Omega}} f\left(\frac{d\boldsymbol{P}}{d\boldsymbol{Q}}\right)d\boldsymbol{Q} = \int_{\boldsymbol{\Omega}} f\left[\frac{dp(x)}{dq(x)}\right]q(x)d\mu(x) \tag{4}$$
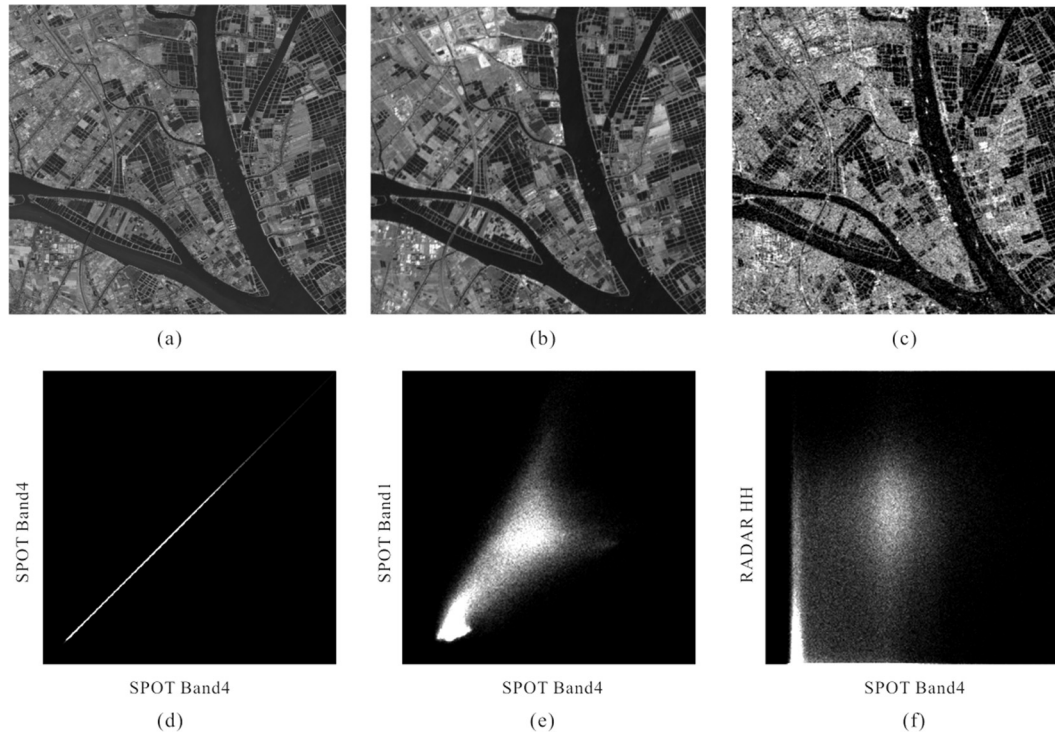
Fig. 1. Different patterns of Joint histograms of different image pairs. (a) SPOT band 1 image; (b) SPOT band 4 image; (c) PalSAR HH polarization image; (d) Joint histogram of two identical SPOT band 4 images; (e) Joint histogram of SPOT band 4 and SPOT band1 images; (f) Joint histogram of SPOT band 4 and PalSAR HH polarization images.

In this definition we have a convention that

$$f\left[\frac{dp(x)}{dq(x)}\right]q(x) = \begin{cases} 0 & \text{if } p(x) = 0, q(x) = 0 \\ p(x)\lim_{x\to\infty}\frac{f(x)}{x} & \text{if } p(x) > 0, q(x) = 0 \end{cases} \tag{5}$$

According to the definitions, f-divergence is naturally a mathematical expectation of the odd ratio given by distributions $\mathbf{P}$ and $\mathbf{Q}$, and weighted by function f. With regard to the special case where $\mathbf{P}$ is identical to $\mathbf{Q}$, we have f-divergence $D_f(\mathbf{P}\|\mathbf{Q}) = 0$, which is the minimum value indicating the closest distance from $\mathbf{P}$ to $\mathbf{Q}$. $D_f(\mathbf{P}\|\mathbf{Q})$ tends to increase as the disparity between $\mathbf{P}$ and $\mathbf{Q}$ increases.

Although f-divergence is often regarded as a "distance" between two distributions, it is, in fact, not symmetric. In other words, the f-divergence from $\mathbf{P}$ to $\mathbf{Q}$ does not necessarily correspond with the f-divergence from $\mathbf{Q}$ to $\mathbf{P}$. Statistically, in $D_f(\mathbf{P}\|\mathbf{Q})$, distribution $\mathbf{Q}$ is often regarded as a model or theoretical distribution of $\mathbf{P}$, whereas $\mathbf{P}$ always represents the distribution of an observation sample of $\mathbf{Q}$. Thus, $D_f(\mathbf{P}\|\mathbf{Q})$ is a measure that measures the unidirectional disparity from the observation result to the reference model. In the context of remotely sensed image registration, $D_f(\mathbf{P}\|\mathbf{Q})$ is favored to be adopted to mark the proximity (alignment) from the floating image to the reference one.

3.2. Mutual information (Kullback-Leibler divergence)

With a specific kernel function $f(\cdot)$, the f-divergence will specialize into different divergences that constitute the f-divergence family. Kullback-Leibler divergence (KL divergence, also known as information divergence or relative entropy) is one of these divergences (Kullback and Leibler, 1951) with kernel function $f(x) = x\log(x)$. Thus, KL-divergence is defined as:

$$D_{KL}(\mathbf{P}\|\mathbf{Q}) = \int_{\Omega} p(x) \log\left[\frac{p(x)}{q(x)}\right] dx \tag{6}$$

KL divergence was first studied by Kullback and Leibler in the 1950s as a directed divergence from distribution $\mathbf{P}$ to distribution $\mathbf{Q}$

(Kullback and Leibler, 1951). The logarithmic-like function $f(\cdot)$ is analogous to the kernel function of the Shannon entropy:

$$H(\mathbf{P}) = -\int p(x) \log[p(x)] dx \tag{7}$$

Shannon entropy $H(\mathbf{P})$ measures the self-information of a signal $\mathbf{P}$ while KL divergence $D_{KL}(\mathbf{P}\|\mathbf{Q})$ indicates the relative entropy from signal $\mathbf{P}$ to signal $\mathbf{Q}$. **Mutual information** is a special case of KL divergence: it can only be applied to a situation where distribution $\mathbf{P}$ is a joint distribution $\mathbf{P}_{XY}$ and $\mathbf{Q}$ is the product of two marginal distributions $\mathbf{P}_X \times \mathbf{P}_Y$, which is defined as

$$\begin{aligned} MI(\mathbf{P}_X; \mathbf{P}_Y) &= D_{KL}(\mathbf{P}_{XY}\|\mathbf{P}_X \times \mathbf{P}_Y) \\ &= \iint p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x) \times p_Y(y)} dxdy \end{aligned} \tag{8}$$

In information theory, mutual information denotes the common information shared between $\mathbf{P}_X$ and $\mathbf{P}_Y$ by measuring the KL-divergence from the joint distribution $\mathbf{P}_{XY}$ to the product of the two marginal distributions $\mathbf{P}_X$ and $\mathbf{P}_Y$ (the product of two marginal distributions is actually the hypothetical joint distribution if $\mathbf{X}$ and $\mathbf{Y}$ are independent). In other words, mutual information quantifies the dependence between the joint distribution of $\mathbf{X}$ and $\mathbf{Y}$ and what the joint distribution would be if $\mathbf{X}$ and $\mathbf{Y}$ are independent. It is the same case to any divergences between $\mathbf{P}_{XY}$ and $\mathbf{P}_X \cdot \mathbf{P}_Y$. The higher this divergence, the more dependent and the more similar they are. Note that, despite slight difference between the "KL-divergence" and "mutual information" in the field of probability theory, they refer to the same measure in the context of image registration. In the rest of the paper, the "KL-divergence" and "mutual information" are identical and refer to the measure defined in Eq. (8). In the mathematical analysis, we will use the "KL-divergence" to keep it mathematically rigorous, but use the "mutual information" in the experimental description to keep it consistent with current literature.

In image registration, mutual information varies if the overlap region of the image pair changes, which can be manifested through the distribution of intensity pairs in the joint histogram. If the floating image is perfectly aligned to the reference image, the strongest dependency among all intensity pairs in these two images exists. In this case, all intensity pairs aggregate and mutual information obtains its maximum value. In all other cases, dependency between the image pair becomes weaker. Consequently, intensity pairs in the join histogram become dispersive and mutual information will consequently decline.

3.3. Limitation of mutual information in image registration

So far, mutual information and its modified forms have been used in remotely sensed image registration and have achieved satisfactory accuracy in some multimodal image registrations (Inglada and Giros, 2004; Inglada et al., 2007; Kern and Pattichis, 2007). In these models, the probability distributions of an image pair are evaluated according to Eq. (9), where $\boldsymbol{P_{XY}}$ is estimated by the joint histogram of the image pair; and $\boldsymbol{P_X}$ and $\boldsymbol{P_Y}$ are estimated by the histogram of the overlap region in the two images respectively, i.e.,

$$\begin{aligned} \boldsymbol{P_{XY}} &= \frac{1}{\sum_{ij} c_{ij}} \mathbf{JH} \\ \boldsymbol{P_X} &= \sum_Y \boldsymbol{P_{XY}} \\ \boldsymbol{P_Y} &= \sum_X \boldsymbol{P_{XY}} \end{aligned} \tag{9}$$

where $\mathbf{JH}$ is the joint histogram of the image pair, and c_{ij} is the element in the $\mathbf{JH}$ matrix. The search methodology in the registration process is equivalent to a mathematical optimization problem. The optimal solution corresponds with the best transformation parameters that lead to the maximum of the similarity measure, which yields the optimum registration of the image pair. Thus, the similarity measure must meet the following **presupposition**: the floating image perfectly aligns to the reference image *if and only if* the similarity measure obtains its maximum within the feasible search space. However, mutual information may fail to yield the optimum registration when the image pair has a small overlap region.

In remotely sensed images, the spectral reflectance in a pixel is usually related to its neighbor's because ground objects are arranged by orderliness and systematic concentration rather than randomness. Thus, the mutual information of two aligned images is relatively high. When shifting away from each other, the mutual information of two misaligned images begins to decline. Theoretically, if the shift distance is greater than a critical value,[2] the correlation between two pixels will be negligible (Clark and Harper, 2000), which implies the mutual information is negligible as well. Previous studies (Li et al., 2008; Liu et al., 2007; Maes et al., 1997; Viola and Wells, 1997) have shown that mutual information declines if the perfectly aligned image pair moves away from each other (Fig. 2-a). However, we found that, if the image pair continues to shift in a large scale and exceed the above-mentioned value, resulting in a small overlap ratio, mutual information will become unstable and exceed the desired one that indicates the optimum transformation parameters (Fig. 2-b). Consequently, previous models based on finding the maximum mutual information value will fail to identify the optimum registration.

To analyze the variation of mutual information and its connection to overlap ratio, we start with a simplified registration case

where the reference image and the floating image are identical and perfectly aligned. Under this circumstance, all the intensity pairs in the joint histogram will aggregate on the diagonal as the two images are perfectly aligned, and mutual information will obtain its maximum value. We have

$$\boldsymbol{P_{XY}} = \frac{1}{\sum_{ij} c_{ij}} \mathbf{JH} = \begin{cases} p_{ij} & i = j \\ 0 & i \neq j \end{cases} \tag{10}$$

$$\begin{aligned} \boldsymbol{P_X} &= \sum_Y \boldsymbol{P_{XY}} = \begin{bmatrix} p_{11} & p_{22} & \cdots & p_{ii} & \cdots & p_{MM} \end{bmatrix} \\ \boldsymbol{P_Y} &= \sum_X \boldsymbol{P_{XY}} = \begin{bmatrix} p_{11} & p_{22} & \cdots & p_{ii} & \cdots & p_{MM} \end{bmatrix}^T \end{aligned} \tag{11}$$

Then, mutual information of a perfectly aligned image pair is estimated according to the definition in Eq. (8). For simplicity, the images are specialized into discrete distributions:

$$\begin{aligned} MI^*(\mathbf{I}_R; \mathbf{I}_F) &= D_{KL}(\boldsymbol{P_{XY}} || \boldsymbol{P_X} \times \boldsymbol{P_Y}) \\ &= \sum_{XY} p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x) \times p_Y(y)} = -\sum_i p_{ii} \log(p_{ii}) \\ &= H(\mathbf{I}_R) = H(\mathbf{I}_F) \end{aligned} \tag{12}$$

In this equation, $MI^*(\mathbf{I}_R; \mathbf{I}_F)$ is the mutual information of the image pair $\mathbf{I}_F$ and $\mathbf{I}_R$ if the two images are perfectly aligned, and $H(\mathbf{I})$ is the Shannon entropy of the image $\mathbf{I}$. According to Eq. (12), the mutual information of these two identical images is equal to the Shannon entropy of the image itself. As regards to a remotely sensed image, the Shannon entropy of the entire image is always greater than that of a subset image.[3] That is,

$$\begin{aligned} H(\mathbf{I}'_R) &\leqslant H(\mathbf{I}_R) \\ H(\mathbf{I}'_F) &\leqslant H(\mathbf{I}_F) \end{aligned} \tag{13}$$

where $\mathbf{I}'_R$ and $\mathbf{I}'_F$ are the overlap regions of the reference image and the floating image respectively. According to information theory, the mutual information of two signals is always smaller than the minimum of their Shannon entropy:

$$MI(\boldsymbol{P}; \boldsymbol{Q}) \leqslant \min\{H(\boldsymbol{P}), H(\boldsymbol{Q})\} \tag{14}$$

Therefore, during the registration process, we have

$$\begin{aligned} MI(\mathbf{I}'_\mathbf{R}; \mathbf{I}'_\mathbf{F}) &\leqslant \min\{H(\mathbf{I}'_\mathbf{R}), H(\mathbf{I}'_\mathbf{F})\} < \min\{H(\mathbf{I}_\mathbf{R}), H(\mathbf{I}_\mathbf{F})\} \\ &= MI^*(\mathbf{I}_\mathbf{R}, \mathbf{I}_\mathbf{F}) \end{aligned} \tag{15}$$

in which $MI(\mathbf{I}'_R; \mathbf{I}'_F)$ is the mutual information value obtained if the transformed image $\mathbf{I}_F(\boldsymbol{G_\mu} \circ (\boldsymbol{X}'))$ and the reference image $\mathbf{I}_R(\boldsymbol{X})$ have their overlap regions $\mathbf{I}'_F$ and $\mathbf{I}'_R$.

According to Eq. (15), even though the mutual information value $MI(\mathbf{I}'_R; \mathbf{I}'_F)$ may drift up and down due to the small overlap region, $MI(\mathbf{I}'_R; \mathbf{I}'_F)$ will not exceed the desired $MI^*(\mathbf{I}_R; \mathbf{I}_F)$ obtained when the two identical images are perfectly aligned. It means that, in this simplified case, as long as we find the maximum of mutual information throughout the optimization process, we can always obtain the desired transformation $\boldsymbol{T}^*$ to perfectly align the floating image to the reference image, without being trapped into transformation parameters that lead to a small overlap region.

However, multimodal image registration does not offer the same result. Mutual information may exceed the desired one when there is small overlap region in common, which is more likely to happen if these two images are inherently different or are deteriorated by noises. With regard to those multimodal cases, one-to-one

---

[2] In geostatistics, it is referred to as the range at which the semi-variogram (or semi-variogram component) reaches the sill value. Presumably, autocorrelation is essentially zero beyond the range.

[3] Under normal circumstances, we have an assumption that the histogram of one image satisfies a certain distribution according to the *Bernouli large numbers law*, and the information of one image can be depicted by the Shannon Entropy of the very distribution. Thus, the information containing one image is always greater than the information containing in the subset one.
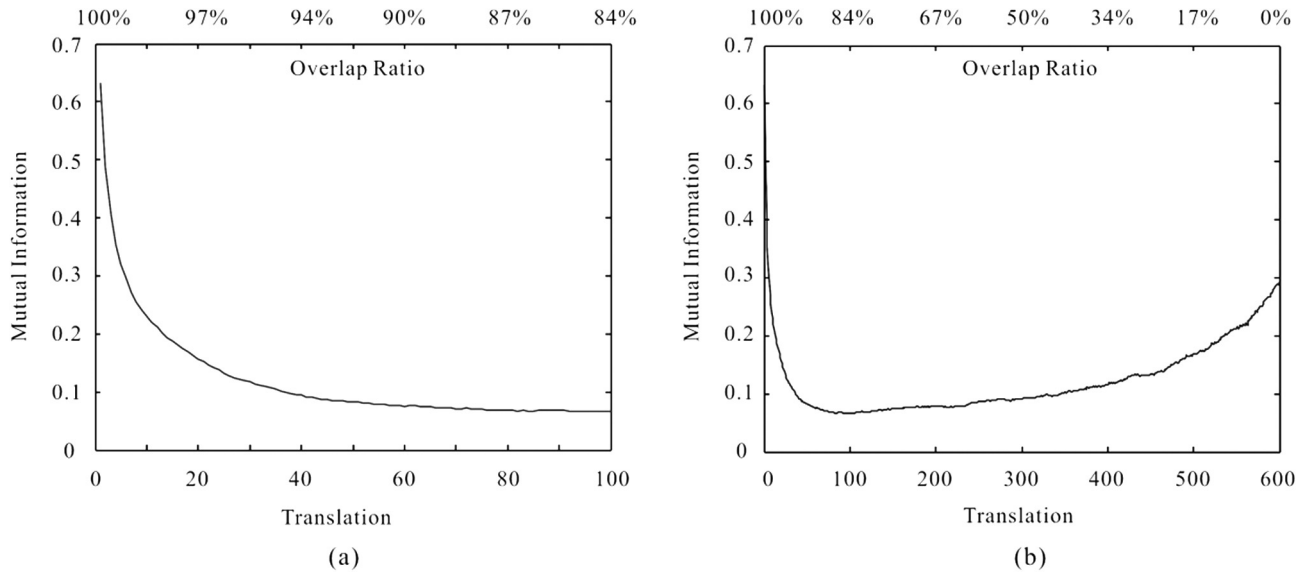
Fig. 2. Mutual information tends to (a) decline when the images of the pre-aligned image pair (SPOT band 1 and SPOT band 4 images) move away from each other along the horizontal direction for small displacements (within 100 pixels), and (b) tends to rebound for the larger displacements (within 600 pixels).

mapping no longer exists between the intensities of the two images. Intensity pairs populating in the joint histogram are not aggregated in the diagonal, which results in

$$MI^*(\mathbf{I}_R; \mathbf{I}_F) < \min\{H(\mathbf{I}_R), H(\mathbf{I}_F)\} \tag{16}$$

In this case, we cannot determine whether $MI^*(\mathbf{I}_R; \mathbf{I}_F)$ is greater than $MI(\mathbf{I}'_R; \mathbf{I}'_F)$ or not. Moreover, $MI^*(\mathbf{I}_R; \mathbf{I}_F)$ from Eq. (16) is much smaller than that from Eq. (12) due to the inherent differences between the image pair. As the multimodal image pair moves away from each other, the mutual information of the overlap images initially declines because of the reduction of autocorrelation, and then rebounds because of the unexpected local similarity of the overlap region. With a smaller overlap region, only small intensity intervals in both images are utilized to compose the joint histogram. Intensity pairs in the joint histogram will thin out and aggregate around specific regions, which results in the instability of mutual information. If we consider an extreme case where there are only a few intensity pairs left in the joint histogram, intensity pairs will "aggregate" to a clustering center. The dependency of these intensity pairs (but not the dependence of the image pair) will be significant, which will lead to a very high value of mutual information. Thus, Eq. (15) is not always correct for all mutual information value $MI(\mathbf{I}'_R; \mathbf{I}'_F) \in (0, \min\{H'(\mathbf{I}_R), H'(\mathbf{I}_F)\})$ obtained during the multimodal image registration process. If the overlap ratio decreases to a critical value, the mutual information $MI(\mathbf{I}'_R; \mathbf{I}'_F)$ will exceed the optimal $MI^*(\mathbf{I}_R; \mathbf{I}_F)$ that is supposed to be the maximum[4] (Fig. 3). Therefore, in practice, multimodal image registration based on mutual information should take the overlap ratio into consideration. The variation of mutual information illustrated in Fig. 3 will lead to an incorrect registration result when we directly apply a search algorithm (gradient descent algorithm, genetic algorithm, etc.) to optimize the mutual information. There exists a particular search space around the optimum transformation parameters (referred to as "*feasible search space*" hereinafter) only within which

the mutual information value obtained at the optimum transformation parameters is the highest. Thus, the search space of transformation parameters has to be restricted to guarantee the presupposition that mutual information obtains its maximum if and only if the image pair is perfectly aligned.

4. Similarity measure based on Jeffrey's divergence

According to the previous analysis, multimodal image registration based on mutual information may yield an unreliable registration outcome because of the small image overlap region. To mitigate this problem, we propose to use the symmetrical form of mutual information, namely Jeffrey's divergence, as the similarity measure for multimodal image registration.

4.1. Definition of Jeffrey's divergence

Jeffrey's divergence (abbreviated as J-divergence), was first proposed by Jeffreys (1939, 1946) as the "divergence" between two distributions $\mathbf{P}$ and $\mathbf{Q}$, defined as

$$
\begin{aligned}
D_J(\mathbf{P}, \mathbf{Q}) &= D_{KL}(\mathbf{P}||\mathbf{Q}) + D_{KL}(\mathbf{Q}||\mathbf{P}) \\
&= \int_\Omega [p(x) - q(x)] \log\left[\frac{p(x)}{q(x)}\right] dx
\end{aligned} \tag{17}
$$

The definition implies that J-divergence is actually a symmetrical form of KL-divergence by adding $D_{KL}(\mathbf{P}||\mathbf{Q})$ and $D_{KL}(\mathbf{Q}||\mathbf{P})$. According to the definition of f-divergence given in Eq. (4), J-divergence can also be regarded as a specific case of the f-divergence family, with the kernel function

$$f(x) = (x - 1)\log(x) \tag{18}$$

That is,

$$D_J(\mathbf{P}, \mathbf{Q}) = \int_\Omega f\left[\frac{p(x)}{q(x)}\right] q(x) dx = \int_\Omega [p(x) - q(x)] \log\left[\frac{p(x)}{q(x)}\right] dx \tag{19}$$

In the context of image registration, J-divergence quantifies the discrimination of an image pair by measuring the "distance" bi-directionally between the joint histogram and the product marginal histograms of two images. Analogous to the mutual information of an image pair, we define the J-divergence between two images as
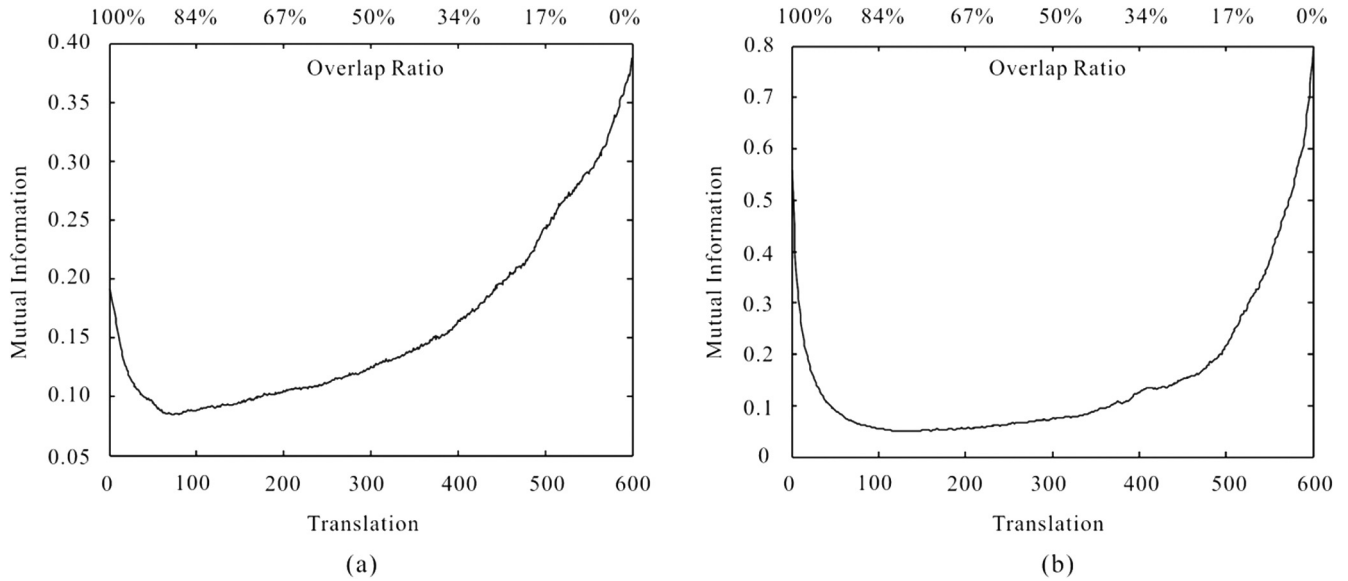
---

[4] The accuracy of image registration greatly depends on the interpolation method used in the transformation process. Since we are not aiming at achieving sub-pixel accuracy, but at analyzing its robustness to overlap ratio and noise interference, accurate interpolation models (e.g., PVE, GPVE) do not contribute to the result, but introduce extra computational cost. Therefore, we only use the simplest one, namely, nearest neighbor interpolation.

Fig. 3. Variations of the mutual information as the images of the multimodal image pair move away from each other along the horizontal direction. The value of mutual information is much higher when the image pair is perfectly aligned than when there is a great horizontal malposition. (a) SPOT band 4 and PalSAR HH; (b) SPOT band 4 and Landsat TM band 5.

$$D_J(\mathbf{I}_R, \mathbf{I}_F) = D_J(\boldsymbol{P}_{XY}, \boldsymbol{P}_X \times \boldsymbol{P}_Y)$$
$$= \iint [p_{XY}(x,y) - p_X(x) \cdot p_Y(y)] \log \frac{p_{XY}(x,y)}{p_X(x) \cdot p_Y(y)} dx dy \quad (20)$$

where similar to mutual information, $\boldsymbol{P}_{XY}$ is the joint probability distribution estimated from the joint histogram, and $\boldsymbol{P}_X$ and $\boldsymbol{P}_Y$ are the marginal probability distributions estimated from the marginal histogram of the image.

According to the definition of J-divergence, $D_J(\mathbf{I}_R, \mathbf{I}_F)$ can also be written as:

$$D_J(\mathbf{I}_R, \mathbf{I}_F) = D_J(\boldsymbol{P}_{XY}, \boldsymbol{P}_X \times \boldsymbol{P}_Y)$$
$$= D_{KL}(\boldsymbol{P}_{XY} \| \boldsymbol{P}_X \times \boldsymbol{P}_Y) + D_{KL}(\boldsymbol{P}_X \times \boldsymbol{P}_Y \| \boldsymbol{P}_{XY}) \quad (21)$$

The first term $D_{KL}(\boldsymbol{P}_{XY} \| \boldsymbol{P}_X \times \boldsymbol{P}_Y)$ is the KL-divergence from $\boldsymbol{P}_{XY}$ to $\boldsymbol{P}_X \times \boldsymbol{P}_Y$, namely mutual information of two images. The mathematical expression of the second term $D_{KL}(\boldsymbol{P}_X \times \boldsymbol{P}_Y \| \boldsymbol{P}_{XY})$ looks like the KL-divergence from $\boldsymbol{P}_X \times \boldsymbol{P}_Y$ to $\boldsymbol{P}_{XY}$, but in fact, it does not strictly obey the definition of KL-divergence in the registration context. KL-divergence $D_{KL}(\boldsymbol{P} \| \boldsymbol{Q})$ is only defined if $\boldsymbol{P}$ is continuous with respect to $\boldsymbol{Q}$. In other words, the definition only exits if $\boldsymbol{Q}(x) = 0$ implies $\boldsymbol{P}(x) = 0$ for all x. In the context of image registration, P_{XY} is estimated from the joint histogram of two images; $\boldsymbol{P}_X$ and $\boldsymbol{P}_Y$ are estimated by summing $\boldsymbol{P}_{XY}$ in two different directions (see Eq. (9)). As a result, $\boldsymbol{P}_{XY} = 0$ does not imply $\boldsymbol{P}_X \times \boldsymbol{P}_Y = 0$ for all (x,y) (but $\boldsymbol{P}_X \times \boldsymbol{P}_Y = 0$ does imply $\boldsymbol{P}_{XY} = 0$ for all (x,y)), which does not satisfy the absolute continuity in the KL- divergence definition. Since the second term of J-divergence is not a strictly defined KL-divergence, the non-negativity property is not always preserved. Nevertheless, for convenience, we still use the denotation $D_{KL}(\boldsymbol{P}_X \times \boldsymbol{P}_Y \| \boldsymbol{P}_{XY})$ to refer to the dependence from $\boldsymbol{P}_X \times \boldsymbol{P}_Y$ to $\boldsymbol{P}_{XY}$ if necessary.

4.2. Theoretical analysis

When compare the definitions in formula (8) and (20), the definitions of KL-divergence and J-divergence are quite similar but slightly different in the summation terms, and the difference of the summation terms results in the different performances between J-divergence and mutual information. In the definition

of J-divergence, regardless of whether $p_{XY}(x,y)$ is greater than $p_X(x) \cdot p_Y(y)$ or not, we always have a non-negative summation term:

$$[p_{XY}(x,y) - p_X(x) \cdot p_Y(y)] \log \left[\frac{p_{XY}(x,y)}{p_X(x) \cdot p_Y(y)} \right] \geqslant 0 \quad (22)$$

That is, the J-divergence of an image pair $D_J(\mathbf{I}_R, \mathbf{I}_F)$ continues to increase monotonously during the accumulation process. However, mutual information is not the same case. When $p_X(x) \cdot p_Y(y)$ is greater than $p_{XY}(x,y)$, the logarithm part becomes negative and yields

$$p_{XY}(x,y) \log \left[\frac{p_{XY}(x,y)}{p_X(x) \cdot p_Y(y)} \right] \leqslant 0 \quad (23)$$

In the joint histogram, probability $p_{XY}(x,y)$ is proportional to the number of intensity pairs populating location (x,y), whereas probability $p_X(x)$ or $p_Y(y)$ is related to the density of intensity pairs around location (x,y). That is, whether $p_{XY}(x,y)$ is greater than $p_X(x) \cdot p_Y(y)$ or not depends on the distribution of intensity pairs at and around location (x,y). In some locations close to the cluster center of intensity pairs (see Fig. 4, the red point), $p_{XY}(x,y)$ is significantly greater than $p_X(x) \cdot p_Y(y)$ because many intensity pairs are in this location. The other (x,y) located at the fringe of the intensity pairs' aggregation zone (see Fig. 4, the blue point) will encounter the situation where $p_{XY}(x,y)$ is smaller than $p_X(x) \cdot p_Y(y)$, which will yield a negative summation term based on the definition of mutual information. Note that no matter which quantitative relation between $p_{XY}(x,y)$ and $p_X(x) \cdot p_Y(y)$, all summation terms will finally yield a non-negative value KL-divergence value.

If the multimodal image pair is aligned perfectly, the intensity pairs in the joint histogram will aggregate around certain regions. For every location (x,y), the summation term in J-divergence meets Eq. (22) and stays non-negative. However, mutual information will encounter several locations that lead to Eq. (23) and will not keep increasing monotonously. If the image pair is significantly different in spectrum/coherency or contaminated by noises, the distribution of intensity pairs in the joint histogram will be more dispersive. More negative summation terms will emerge during the accumulation of mutual information, which is why mutual information declines as random factors interfere with the image pair. Superior
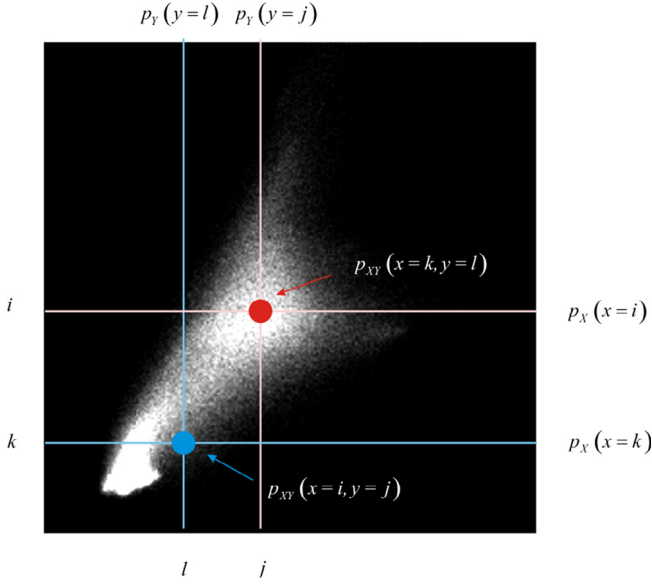
Fig. 4. Joint probability estimation of the intensity-pairs in the joint histogram, and its relationship with intensity-pair distribution around it. The red intensity-pair populates around the cluster center, which will yield out $p_{XY}(x,y) > p_X(x) \cdot p_Y(y)$; while the blue one is located at the fringe of the aggregation zone, which will turn out to be $p_{XY}(x,y) < p_X(x) \cdot p_Y(y)$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to mutual information, the summation term in J-divergence remains non-negative even though the intensity pairs disperse because of the difference of image pair and the noise interference. Thus, compared with mutual information, J-divergence is capable of obtaining a stable and higher value when the multimodal image pair is perfectly aligned.

As analyzed in the Section 3.3, if only a few overlap regions exist between image pairs, then only a small intensity interval in both images will be utilized to compose the joint histogram. The intensity pairs in the joint histogram are so aggregative that a majority of locations satisfy the inequality $p_{XY}(x,y) \geqslant p_X(x) \cdot p_Y(y)$. Consequently, the mutual information obtains a high value that may exceed the desired one. As to J-divergence, the summation term in every location (x, y) is smaller than that of mutual information based on Eq. (24). Δ is the difference between the summation terms of mutual information and J-divergence. It is non-negative as long as $p_{XY}(x,y) \geqslant p_X(x) \cdot p_Y(y)$. Thus, J-divergence obtains a smaller value compared with mutual information if a small overlap region exists.

$$
\begin{aligned}
\Delta &= p_{XY}(x,y) \log \left[\frac{p_{XY}(x,y)}{p_X(x) \cdot p_Y(y)} \right] - [p_{XY}(x,y) - p_X(x) \cdot p_Y(y)] \\
&\quad \times \log \left[\frac{p_{XY}(x,y)}{p_X(x) \cdot p_Y(y)} \right] \\
&= p_X(x) \cdot p_Y(y) \log \left[\frac{p_{XY}(x,y)}{p_X(x) \cdot p_Y(y)} \right] \geqslant 0 \text{ if } p_{XY}(x,y) \\
&\geqslant p_X(x) \cdot p_Y(y)
\end{aligned}
\tag{24}
$$

According to the above analysis, J-divergence obtains a higher value than mutual information if the image pair is perfectly aligned, but a lower value than mutual information if only small overlap region exists between these two images. This condition means that, in the multimodal image registration equivalent to a mathematical optimization problem, J-divergence is capable of providing a larger feasible search to determine the optimum registration. Therefore, image registration models based on J-

divergence is theoretically superior to mutual information in the applicability of registering multimodal images that contains insufficient overlap region.

4.3. Experiments and discussions

The registration model based on J-divergence and mutual information are applied to image pairs that captured from different sensors. According to the deduction in the previous section, registration model based on J-divergence is capable of providing a larger feasible search space. Thus, we can reasonably presume that J-divergence will theoretically be superior to mutual information in multimodal image registration.

4.3.1. Variation of J-divergence in multimodal registration

In multimodal image registration, no direct relationship exists between intensities in the two images. The dependence between the image pair is much weaker compared to those in multi-spectral image pairs. As a result, the mutual information value in the registration model will be lower in the perfectly aligned position and will suffer more peaks in the search surface, contributing to the unreliability of the registration model. With its symmetrical properties, J-divergence will show better performances than mutual information in cases of multimodal image registration. A comparative study on J-divergence and mutual information is conducted through registrations of PalSAR image, Landsat TM image, and DEM data, using SPOT Band 4 as the reference image. The 2-dimensional variations (surface) of mutual information and J-divergence as the floating images move away from the perfectly aligned position are illustrated in Fig. 5. We can see that both mutual information and J-divergence obtain peak values when the image pairs are aligned perfectly (X Translation = 0, Y Translation = 0), but J-divergence can obtain a much higher value than that of mutual information. When the two images move away from each other, both J-divergence and mutual information decline, but the value of J-divergence is relatively lower than mutual information.

A special situation is presented where the floating image moves along the diagonal of the reference image (Fig. 6). In such case, the overlap ratio between the image pair declines most rapidly. We can easily observe the variation and feasible search space of both similarity measures regarding to the affection of overlap ratio. In Fig. 6 that both similarity measures obtain high values at when the image pair is aligned perfectly (Translation = 0). Comparing to mutual information, J-divergence can obtain a value nearly twice as high. When the two images move away from each other, both J-divergence and mutual information decline. As we continue to move the floating image and squeeze the overlap region, J-divergence and mutual information will rise up and then exceed the values obtained in the perfectly aligned position. Nevertheless, J-divergence could obtain an even lower value than mutual information. Rectangles in the figures suggest the feasible search spaces in terms of both horizontal and vertical translations. As to these multimodal image pairs, J-divergence can provide much larger feasible search spaces (the red rectangle), allowing search algorithms to specify the maximum similarity and determine the optimum registration.

4.3.2. Multimodal registration on individual dimension

To examine the applicability of J-divergence in practical registration tasks, we apply the registration model to several multimodal image pairs that are manually misaligned. Fig. 7 presents the reference image (SPOT Band 4) and floating images (ALOS PalSAR HH polarization, Landsat TM Band 5, and DEM data) that were used in the registration experiments. The SPOT Band 4 image is the 4th band of SPOT-5 Level-2A orthorectified product. The ALOS Pal-
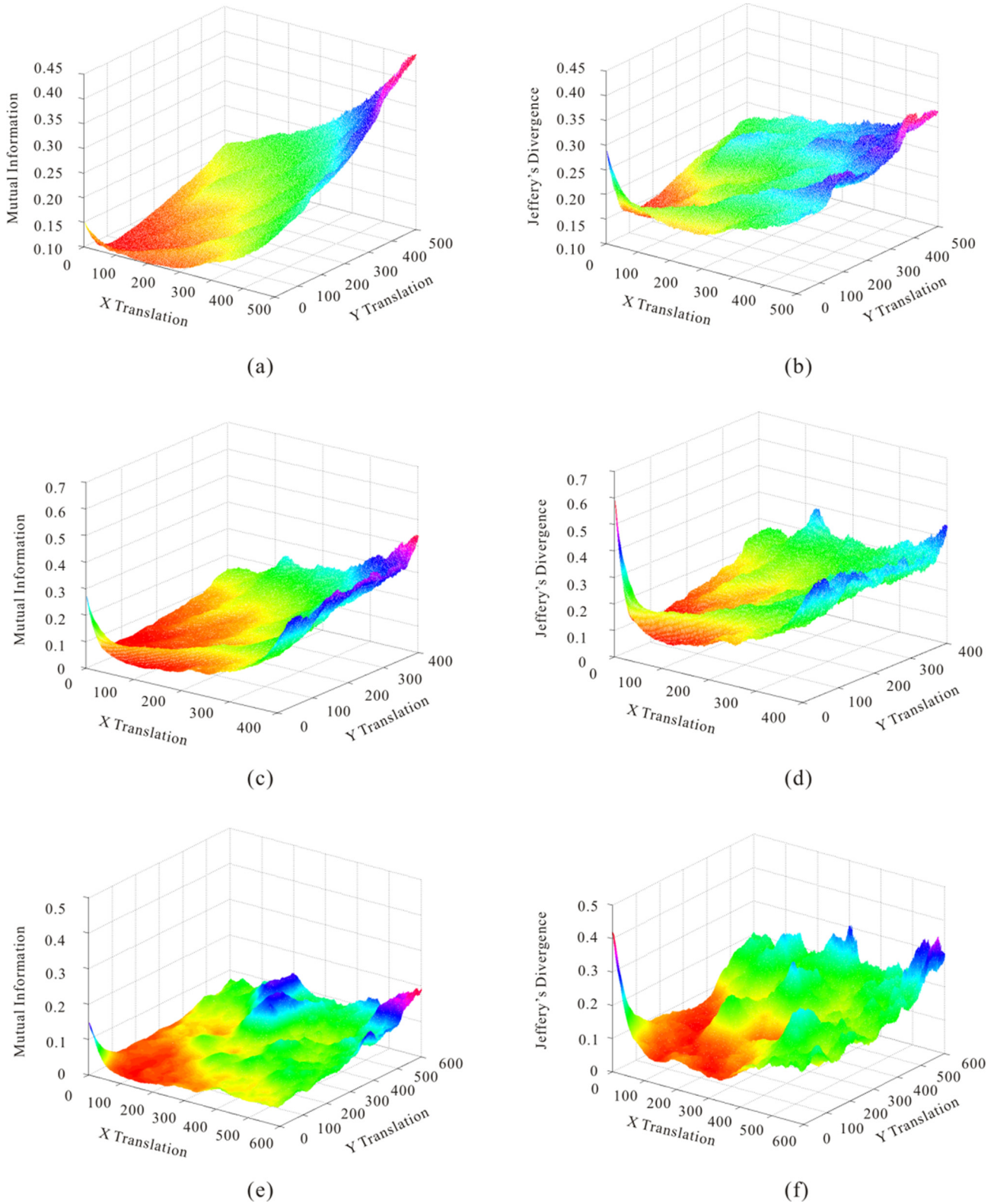
Fig. 5. Two-dimension surfaces of mutual information and Jeffery's divergence when the images of the image pair move away from each other in both horizontal and vertical directions. (a) mutual information of SPOT band 4 and PalSAR HH; (b) Jeffery's divergence of SPOT band 4 and PalSAR HH; (c) mutual information of SPOT band 4 and Landsat TM band 5; (d) Jeffery's divergence of SPOT band 4 and Landsat TM band 5; (e) mutual information of SPOT band 4 and DEM data; (f) Jeffery's divergence of SPOT band 4 and DEM data.

SAR (HH polarization) image was radiometrically and terrain corrected by the data provider (Alaska Satellite Facility, ASF). The Landsat TM image is a Level-1T (orthorectified) Landsat 5 product.

The DEM data was acquired from the Shuttle Radar Topographic Mission (SRTM) at 1 arc-second (~30 m) spatial resolution. Since the topographical variation of the selected region ranges from
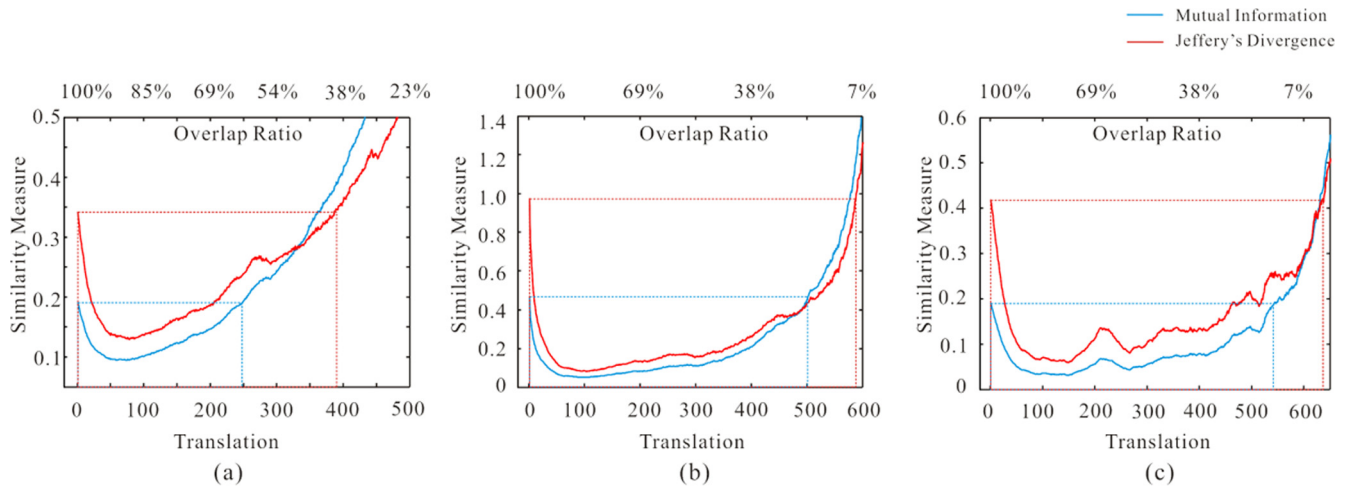
Fig. 6. Variations of Jeffery's divergence and mutual information as the floating images move away along the diagonal of the reference image, with SPOT band 4 as the reference image and PalSAR HH (a), Landsat TM band 5 (b), DEM data (c) as floating images. The red rectangle in each sub-figure reveals the feasible search spaces in these two dimensions using Jeffery's divergernce as the similarity measure while the blue rectangles reveal that of mutual information.
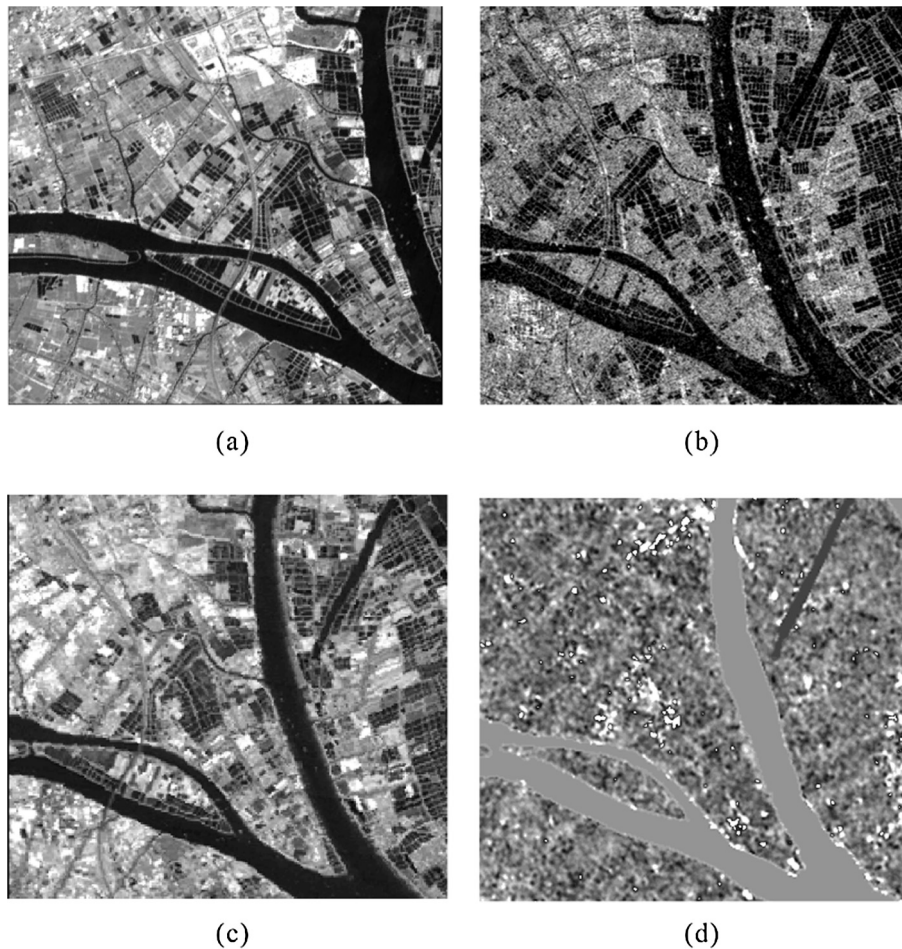

Fig. 7. Multimodal images, which are misaligned, are used as test data to examine the applicability of J-divergence in practical registration tasks, (a) SPOT band 4, (b) PalSAR HH polarization, (c) Landsat TM band 5, (d) DEM data. The geographical extension of the reference image (a) is 200 pixels left of the floating images (b, c and d).

−7 m to 30 m, we limited the ramp color within such range to increase the color contrast. All images were resampled to 10 m resolution and gridded into the same spatial reference. Each image contains 700 ∗ 700 pixels. The geographical extension of the reference image (SPOT Band 4) is 200 pixels left of the floating images.

In the experiment, we moved the misaligned floating images across the reference in the horizontal direction and then examine the variations of the two similarity measure. The translation parameter of horizontal direction is from 0 to 650 pixels (overlap ratio declines from 100% to 8%). Values of Jeffrey's divergence
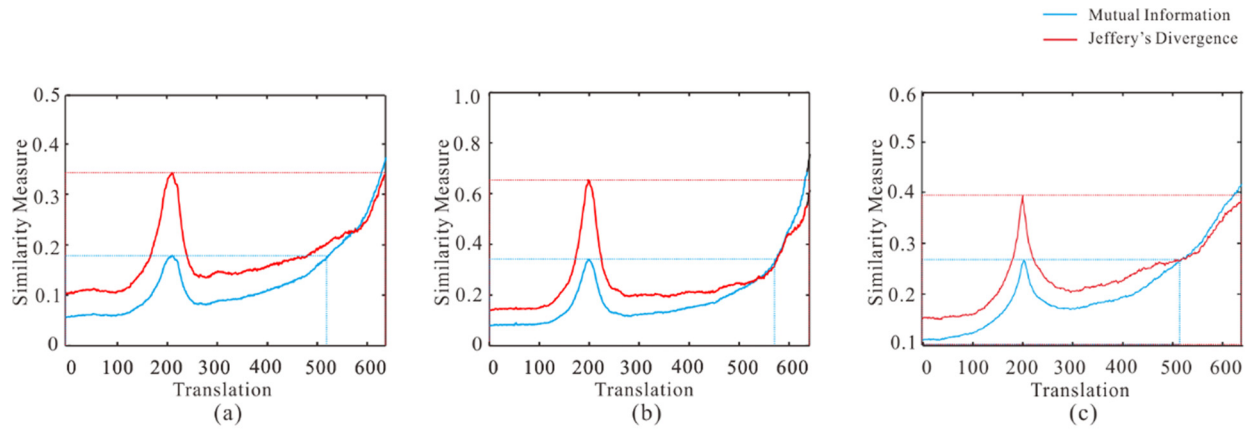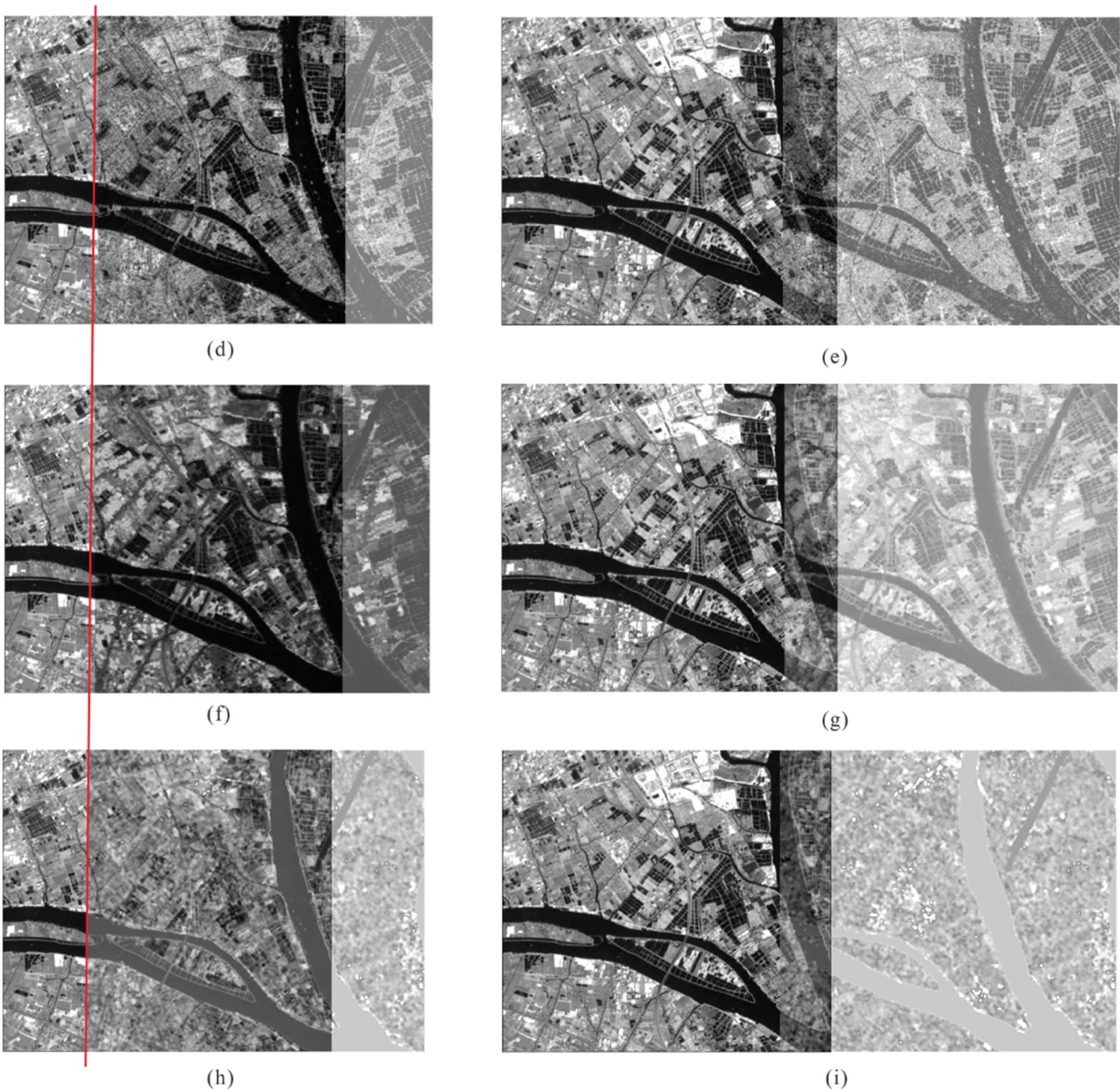
Fig. 8. Performances of mutual information and J-divergence involving misaligned multimodal image pair cases. Variations of the two similarity measures as the images of the misaligned image pair (displacement = 200 pixels) moves across each other along horizontal direction are presented in (a), (b) and (c), with SPOT band 4 as the reference image and PalSAR HH, Landsat TM Band5, DEM data as floating images. The red rectangles in (a), (b) and (c) reveal the feasible search spaces in horizontal direction using Jeffery's divergernce as the similarity measure while the blue rectangles reveal that of mutual information. Sub-image (d–i) are the corresponding registration results within a wide range of searching space, using Jeffrey's divergence (d, f, and h) and mutual information (e, g, and i) as the similarity measures.

and mutual information at each translation pixel are enumerated. A promising similarity measure should give a peak at 200-pixel and lower value even if a small area is overlapped by two images. By identifying the maximum similarity value, a sound search algorithm will find out the correct transformation parameters (at 200-pixel) to register the floating images.

Fig. 8 shows the variations of J-divergence and mutual information in the experiments (Fig. 8-a–c), as well as the registration results corresponding to the highest similarity measures (Fig. 8-d–i). These registration results are all presented by overlapped image pairs, with the upper images set to 50% transparency so that we can directly see the performances of these two similarity measures. In Fig. 11-a–c, both similarity measures obtain peaks at 200 pixels and exhibit rebound at the fringe of the image extent. J-divergence can obtain a much higher value than mutual information as the image pair is registered and a smaller value when the overlap region narrows. The feasible search space of J-divergence in all three cases, indicated by the red rectangle, are much larger than those of mutual information, which are marked by the blue rectangle. Fig. 8-d–i show the registration results when we search the transformation parameters at a large scale (up to 550 pixels), in which Fig. 8-d, f, and h are using J-divergence as the similarity measure and Fig. 11-e, g, and i are using mutual information. Since J-divergence obtains the maximum value at the 200-pixel-displacement in every case, floating images can be correctly aligned to the reference image. However, with mutual information, the value obtained at the 200-pixel-displacement will be exceeded

when we search the parameter space at a large range, for example, 550-pixel-displacement. Therefore, the registration model based on mutual information will incorrectly align the floating images to the reference image at 550-pixel, because the registration model is established under the assumption that the best registration corresponds to the maximum value of the similarity measure.

In practical registration tasks, transformation of the floating image is determined by five parameters that comprise the affine transformation model: translation and scaling in both horizontal and vertical directions as well as rotation. Thus far, we have only explored the translation components. The following experiments will present the performances of J-divergence and mutual information in all five dimensions when registering different multimodal image pairs. Fig. 9 shows the multimodal images that were used in the experiments: SPOT image (2008-11-10), PalSAR image (2009-05-08), TM image (2009-01-02), and DEM data (2009). All of these images cover the same terrestrial scene. SPOT image was the reference image, and the rest will be the floating images. Since the reference and floating images cover the same scene, they will be registered when the transformation parameters are as follows: *translation* $X = 0$, *translation* $Y = 0$, *scaling* $X = 1$, *scaling* $Y = 1$ and *rotation* $= 0$. Sub-figures in Fig. 10 enumerate the variations of J-divergence and mutual information in five dimensions when registering different multimodal image pairs, from which we can compare the feasible search space of these two similarity measures. Within the feasible search space, the registration model will always generate a correct result with a sound search algorithm.
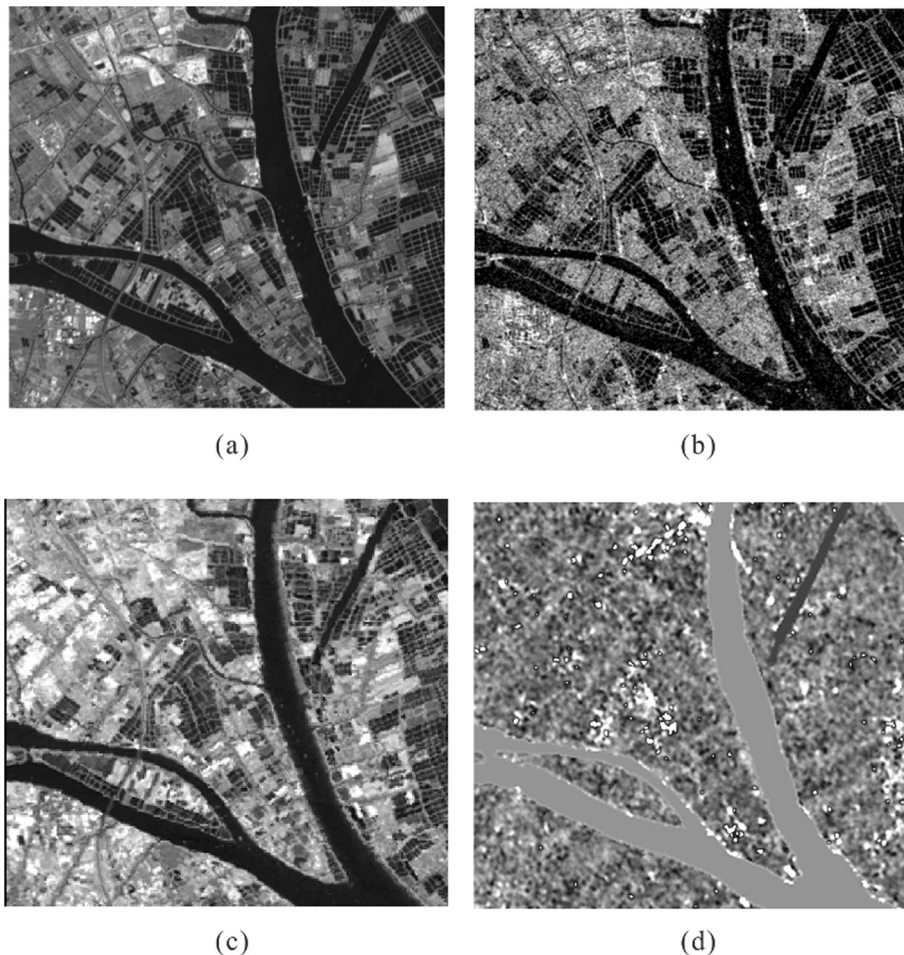

Fig. 9. Multimodal images, which cover the same terrestrial region, were used to examining the performance of J-divergence and mutual information in five dimensions: translation and scaling in both horizontal and vertical directions as well as rotation, (a) SPOT band 4, (b) PalSAR HH polarization, (c) Landsat TM band 5, (d) DEM data.
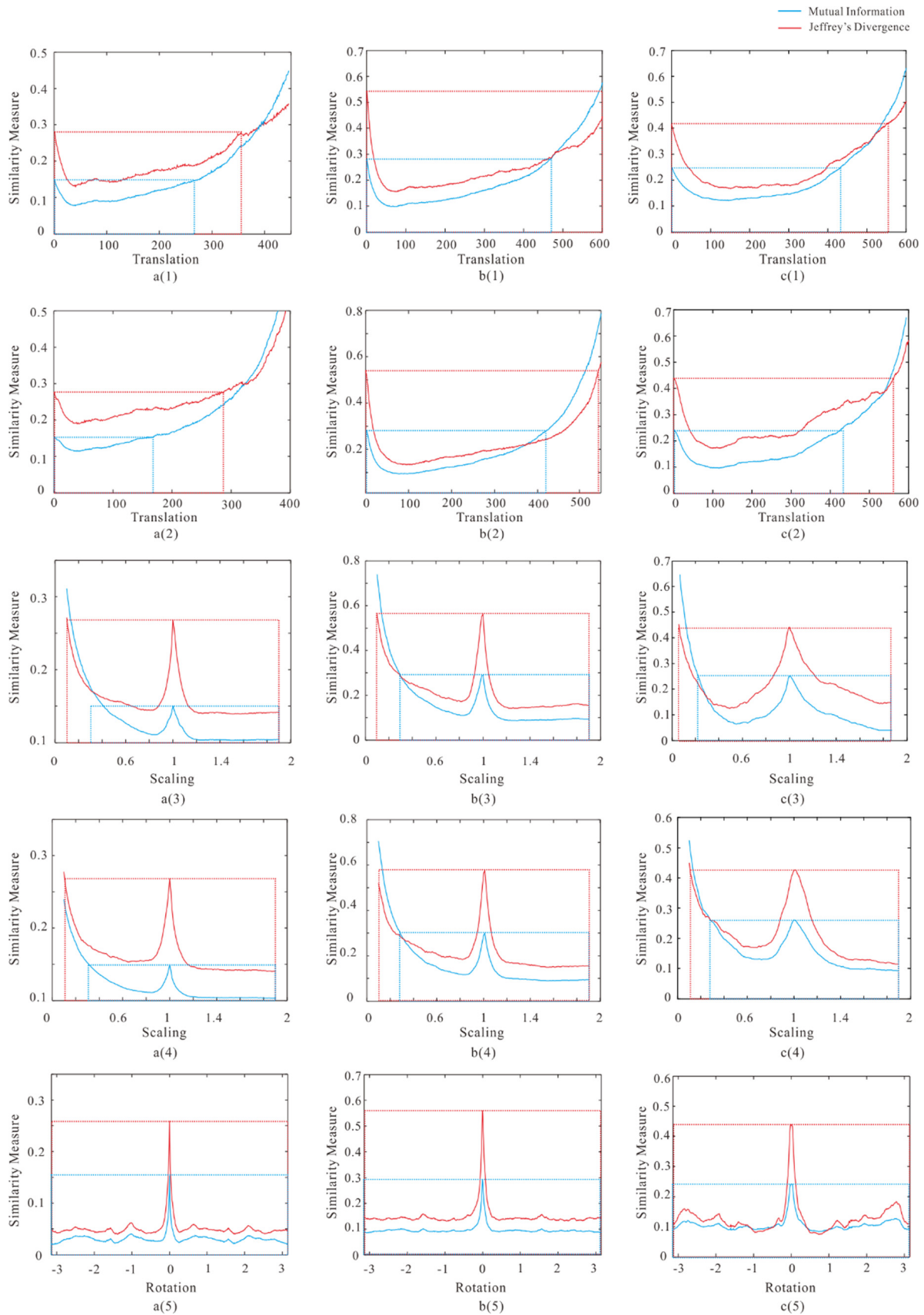
Fig. 10. Variations of Jeffery's divergence and mutual information of an aligned image pair when involving 5 transformation parameters separately, with SPOT band 4 as the reference image and PalSAR HH polarization (a(1)–a(5)), Landsat TM band 5 (b(1)–b(5)) and DEM data (c(1)–c(5)) as floating images. Sub-figure a(1), b(1), c(1) are translation along the horizontal direction; Sub-figure a(2), b(2), c(2) are translation along the vertical direction; Sub-figure a(3), b(3), c(3) are scaling along the horizontal direction; Sub-figure a(4), b(4), c(4) are scaling along the vertical direction; Sub-figure a(5), b(5), c(5) are rotation around the center of the reference image.
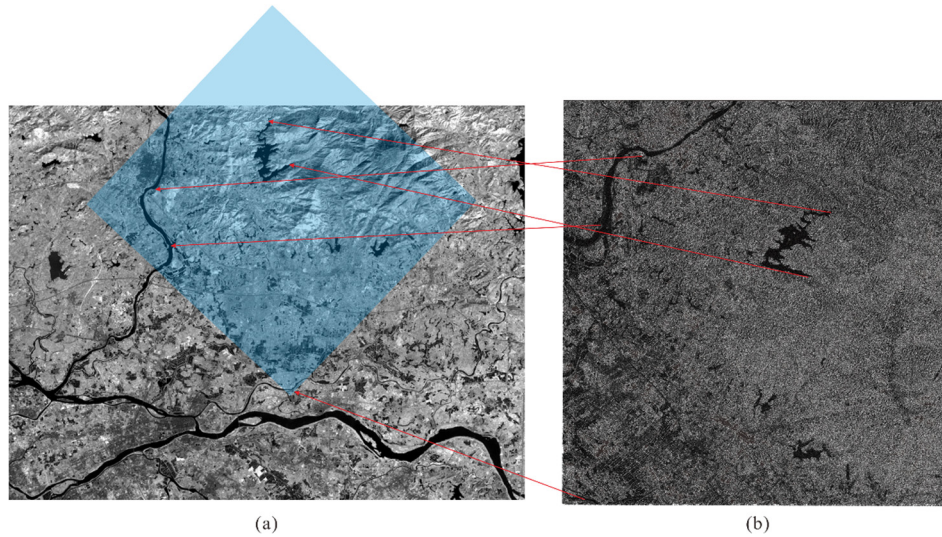
Fig. 11. Data of the comparative registration experiment. (a) Landsat TM band 4 image, (b) ALOS PalSAR image HH polarization image.

Table 1
Examples of divergence family and their kernel functions.

Divergence	Kernel function $f(\cdot)$
Kullback-Leiber (KL-divergence, mutual information)	$x \times \log(x)$
Jeffrey's divergence (J-divergence)	$(x-1) \times \log(x)$
χ^2-divergence	$\frac{1}{2}(x-1)^2$
Lin K-divergence	$x \times \log\left(\frac{2x}{1+x}\right)$
Kolmogorov distance	$\frac{1}{2}|x-1|$

Having a larger size of the feasible search space is equivalent to the capability to register greatly misaligned (or insufficient scene overlap) image pair.

The sub-figures in Fig. 10 show that both similarity measures can obtain a peak at the desired transformation parameters. As to the translation components, sub-figures a (1), a (2), a (3) and b (1), b (2), b (3) show that the variations are similar to those in Fig. 11. J-divergence obtains a higher value than mutual information but a lower value with a smaller overlap region. In scaling components, sub-figures c (1), c (2), c (3) and d (1), d (2), d (3) show that both similarity measures obtain very high values when the scaling is close to 0. As scaling becomes lower, the floating image becomes smaller. As scaling becomes close to 0, the floating image will be transformed into an infinitesimal image, which results in a very small overlap region between the reference and floating images. Thus, the value of similarity will become very high. However, when scaling is larger than 1, the overlap region of the
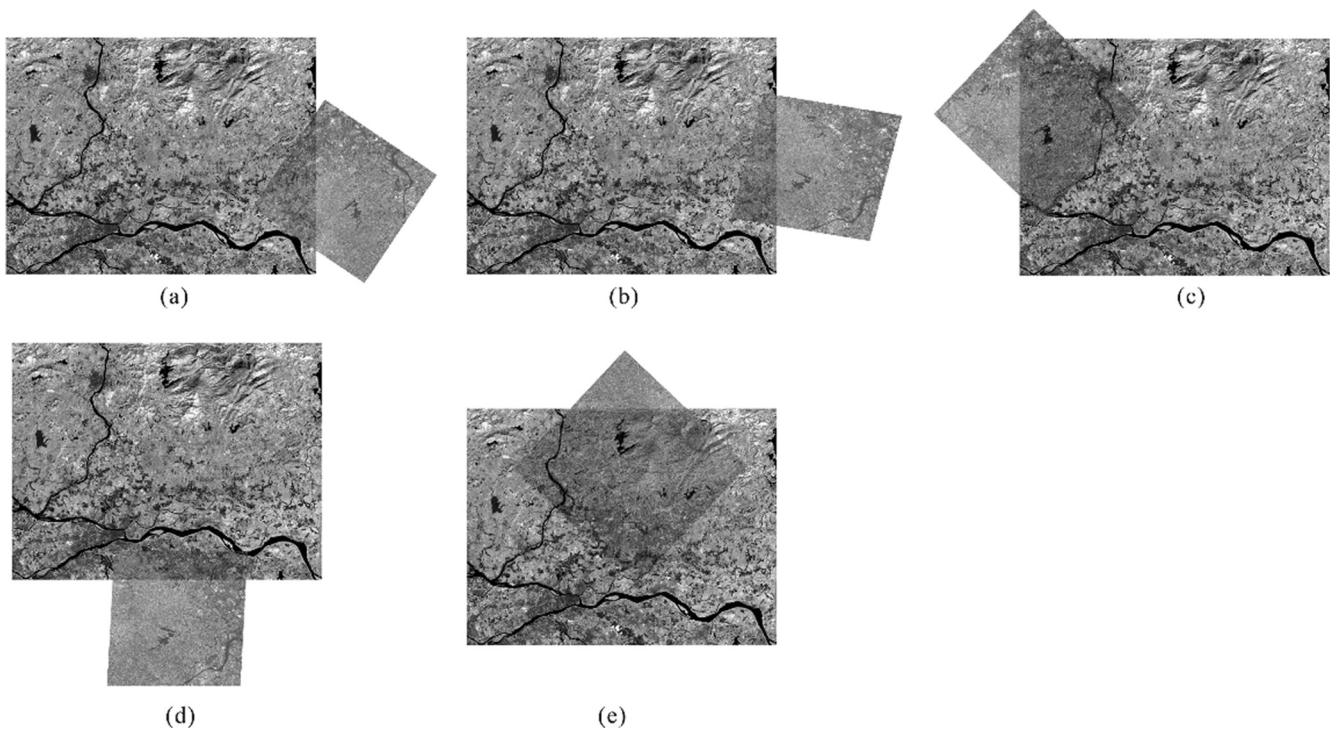

Fig. 12. Registration results by using different similarity measures: (a) mutual information, (b) χ^2-divergence, (c) Lin K-divergence, (d) Kolmogorov distance and (e) J-divergence.
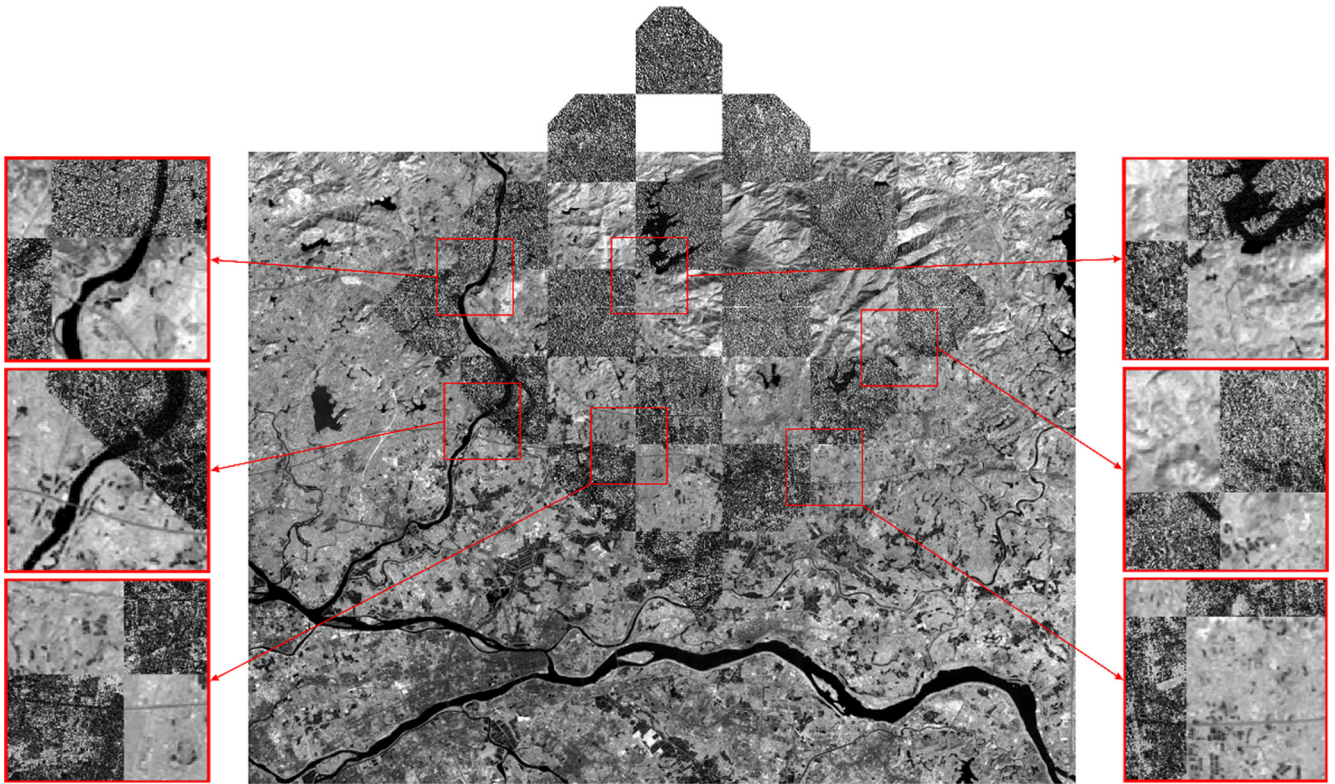
Fig. 13. A chessboard visualization of the registration result based on J-divergence. Six partial enlargements show the favor alignment in the water borders and traffic lines.

Table 2
Registration parameters of five transformation models based on different similarity measures.

Similarity measure	Transformation parameter					Estimated similarity value	Desired[a] similarity value
	Translation X	Translation Y	Scaling X	Scaling Y	Rotation		
Mutual Information	2609	664	0.982	0.979	2.159	0.652	0.373
χ^2-divergence	2770	566	1.131	0.942	3.299	0.591	0.352
Lin K-divergence	−902	−295	0.983	1.011	2.267	0.959	0.865
Kolmogorov distance	998	2189	1.153	1.078	3.135	0.936	0.691
J-divergence	499	−603	0.999	1.009	5.463	0.763	0.763
Manual registration	*498.41*	*−603.52*	*0.999*	*1.000*	*5.465*	*–*	*–*

[a] The "Desired Similarity Value" is the similarity value corresponding to the optimum registration.

Table 3
Comparisons of feasible search space in each dimension and outperformance of J-divergence in providing larger feasible search space, with SPOT being the reference image and PalSAR image as the floating one. The indexes in first column refer to 5 different test image pairs.

Image pair	Similarity measurement	Translation X	Translation Y	Scaling X	Scaling Y	Rotation	Overall improvement
1	MI	(−272,266)	(−186,167)	(0.315,1)	(0.336,1)	(0,2π)	–
	J-divergence	(−362,377)	(−271,288)	(0.104,1)	(0.117,1)	(0,2π)	–
	Improvement	*37.361%*	*58.357%*	*30.803%*	*32.982%*	*0.000%*	*30.492%*
2	MI	(−256,278)	(−215,207)	(0.401,1)	(0.283,1)	(0,2π)	–
	J-divergence	(−353,357)	(−271,288)	(0.212,1)	(0.146,1)	(0,2π)	–
	Improvement	*32.959%*	*57.583%*	*31.553%*	*19.107%*	*0.000%*	*26.838%*
3	MI	(−309,289)	(−241,207)	(0.392,1)	(0.291,1)	(0,2π)	–
	J-divergence	(−386,365)	(−313,277)	(0.145,1)	(0.132,1)	(0,2π)	–
	Improvement	*25.585%*	*31.696%*	*40.625%*	*22.426%*	*0.000%*	*23.279%*
4	MI	(−283,326)	(−276,223)	(0.352,1)	(0.262,1)	(0,2π)	–
	J-divergence	(−359,427)	(−349,368)	(0.139,1)	(0.102,1)	(0,2π)	–
	Improvement	*29.064%*	*43.687%*	*32.870%*	*21.680%*	*0.000%*	*24.559%*
5	MI	(−227,254)	(−236,247)	(0.393,1)	(0.368,1)	(0,2π)	–
	J-divergence	(−314,337)	(−295,326)	(0.219,1)	(0.193,1)	(0,2π)	–
	Improvement	*35.343%*	*28.571%*	*28.666%*	*27.690%*	*0.000%*	*23.379%*

image pair will remain constant (the region covered by the reference image). The similarity between the reference and floating images is irrelevant to overlap interference, but only relevant to the dependence of the image pair. Thus, the similarity measures decrease as the scaling components become larger than 1. As to the rotation component, we rotated the floating image around the center of the reference image. During the rotation, the overlap region of the image pair occupies the majority of the area covered by the reference image. Both similarity measures obtain the highest values at 0-rotaion when we rotate the floating image from $-\pi$ to π (see Fig. 10-a (5), b (5) and c (5)). All sub-figures in Fig. 10 show that J-divergence can obtain a higher value than mutual information at the optimum transformation, but a lower value with a small overlap region. The feasible search space provided by J-divergence is larger than that by mutual information.

4.3.3. Comparative performance of J-divergence in multimodal image registration

So far, we have test the variations of J-divergence in every component of the affine transformation model separately. The J-divergence is capable of providing larger feasible search space in all components except the *Rotation* one. However, in practical multimodal registration tasks, the floating images are transformed synthetically in all components. This is a more complicated process, especially in situations of registering multimodal images covering large area over complex terrain. It is difficult to visualize the 5-dimensional feasible search spaces provided by the similarity measures and intuitively compare their size. In order to examine the applicability of J-divergence in practical registrations of multimodal image pairs, we applied the registration model to the image pair whose scene-covers are partially overlapped. In addition, to justify the superiority of J-divergence, another three divergences: χ^2-divergence, Lin K-divergence and Kolmogorov distance, whose kernel functions are illustrated in Table 1, were also included as similarity measures in the practical registration experiment. To find out the optimum transformation parameters, the model has to search some transformation situations that result in even smaller overlapping than the scene overlapping. In such cases, the similarity value of two images may be very high and probably exceed the one that indicates the optimum registration, leading to an incorrect registration result. Thus the stability of the model greatly depends on the feasible search space provided by the similarity measures.

Fig. 11 presents the reference image (Landsat TM band4, 2009-11-02) and floating one (ALOS PalSAR at HH polarization, 2007-06-24) and their scene overlap region (the blue translucent rectangle). Both the Landsat TM image and ALOS PalSAR image have been orthorectified by the data providers to remove the terrain relief distortions. The TM image was pre-resampled into the same resolution as the ALOS PalSAR image (12.5 m) for the better controlling of the scaling components, since the larger scaling factors will lead to extremely large or small transformed images. The resolution of the TM image is $3221 * 2460$ after resampling, and the resolution of the ALOS PalSAR image is $2428 * 2425$. Both the two images cover complex terrain area with elevation ranging from -4 to 1262 meters. Various surface land cover types, such as cropland, grassland, forest, urban impervious surface and water, etc., are included in this study region.

In the registration process, we have the following configurations to keep the registration model running properly: (1) the overlap region of the two image are set to be not less than 30% while searching the transformation parameters; (2) the scaling components are set to be within the interval of [0.7, 1.5] to avoid an extremely large or small transformed image; (3) with the purpose of eliminating the interference of the search algorithm, we enumer-

ated all the parameter combinations that meet (1) and (2) and estimate their corresponding similarity values. According to the above analysis, the feasible search space provided by J-divergence is significantly larger than that of mutual information. When searching the transformation parameters in a larger range, J-divergence is more likely to identify the correct registration than other similarity measures.

Fig. 12 presents the registration results using five different similarity measures: (a) mutual information, (b) χ^2-divergence, (c) Lin K-divergence, (d) Kolmogorov distance and (e) J-divergence. The transformed ALOS PalSAR images are translucently overlapped on the TM images for better presenting the registration results. All similarity measures except the J-divergence failed to identify the correct transformation parameters since their feasible search space is not large enough. When searching some parameters that lead to a small overlap between two images, these similarities obtain very high values, and such transformation parameters will be incorrectly considered as the optimum registrations. As to J-divergence, even though the overlapping between the image pair decreased to 30%, the estimated value is still lower than that of the desired one corresponding to the perfect alignment. Thus model using J-divergence as the similarity measure can successfully identify the correct transformation parameters according to the maximum of the similarity value. A chessboard visualization of the registration result based on J-divergence is presented in Fig. 13. Six partial enlargements show the favor alignment in the water borders and traffic lines.

Table 2 shows the transformation parameters when the five similarity measures obtain their maximums in the registration process. The bottom row is the transformation parameters of manual registration, which are regarded as the correct transformation parameters of this image pair. The last two columns of the table are the estimated similarity value of the final registrations and the desired similarity values corresponding to the perfect alignment, respectively. We can see that the estimated similarity values of former four measures, including mutual information, χ^2-divergence, Lin K-divergence and Kolmogorov distance, are significantly greater than the desired ones. Thus by finding the maximum value of the similarity measures, they all fail to identify the correct parameters and trap in small overlapping between the image pair. Fortunately, the feasible search space provided by J-divergence is large enough so that the desirable J-divergence value remains the maximum in the registration process. Thus the J-divergence can successfully identify the correct transformation parameters, which are very close to manual registration.

4.3.4. Quantitative assessment

In order to quantitatively assess the feasible search space provided by J-divergence, we presented three tables to show the superiority of J-divergence against mutual information in registering various multimodal image pairs: SPOT and PalSAR (Table 3), SPOT and TM (Table 4), as well as SPOT and DEM (Table 5). Each table presents the experiments of five image pairs which are pre-registered manually. All of these images have been resampled into the same resolution as the SPOT image (10 m),[5] and each image has $650 * 650$ pixels. During the quantitative assessment, we first estimate the J-divergence and mutual information of the registered image pair, which are the similarity values corresponding to the optimum registrations. Then the floating images were gradually deformed along each transformation component (*Translation X*,

---

[5] By resampling all floating images into the same resolution, it will be more convenient to estimate the feasible search space in one dimension while keeping other dimensions constant. Even though the resample process would change the magnitude feasible search space in scaling dimensions, it would not differ the comparison results between J-divergence and mutual information.

Table 4
Comparisons of feasible search space in each dimension and outperformance of J-divergence in providing larger feasible search space, with SPOT being the reference image and Landsat TM image as the floating one. The indexes in first column refer to 5 different test image pairs.

Image pair	Similarity measurement	Translation X	Translation Y	Scaling X	Scaling Y	Rotation	Overall improvement
1	MI	(−412,464)	(−391,423)	(0.291,1)	(0.283,1)	(0,2π)	–
	J-divergence	(−619,630)	(−538,543)	(0.113,1)	(0.097,1)	(0,2π)	–
	Improvement	*42.580%*	*32.801%*	*25.106%*	*25.941%*	*0.000%*	*24.434%*
2	MI	(−453,448)	(−408,461)	(0.269,1)	(0.225,1)	(0,2π)	-
	J-divergence	(−601,595)	(−519,522)	(0.106,1)	(0.113,1)	(0,2π)	-
	Improvement	*32.741%*	*19.793%*	*22.298%*	*14.452%*	*0.000%*	*17.353%*
3	MI	(−419,425)	(−362,450)	(0.304,1)	(0.267,1)	(0,2π)	–
	J-divergence	(−591,611)	(−481,579)	(0.172,1)	(0.126,1)	(0,2π)	–
	Improvement	*42.417%*	*30.542%*	*18.966%*	*19.236%*	*0.000%*	*21.403%*
4	MI	(−436,419)	(−424,461)	(0.293,1)	(0.225,1)	(0,2π)	–
	J-divergence	(−572,553)	(−596,603)	(0.154,1)	(0.107,1)	(0,2π)	–
	Improvement	*31.579%*	*35.480%*	*19.661%*	*15.226%*	*0.000%*	*19.705%*
5	MI	(−422,431)	(−434,467)	(0.241,1)	(0.329,1)	(0,2π)	–
	J-divergence	(−598,615)	(−573,553)	(0.119,1)	(0.163,1)	(0,2π)	–
	Improvement	*42.204%*	*24.972%*	*16.074%*	*24.739%*	*0.000%*	*20.807%*

Table 5
Comparisons of feasible search space in each dimension and outperformance of J-divergence in providing larger feasible search space, with SPOT being the reference images and DEM data as the floating one. The indexes in first column refer to 5 different test image pairs.

Image pair	Similarity measurement	Translation X	Translation Y	Scaling X	Scaling Y	Rotation	Overall improvement
1	MI	(−416,430)	(−431,426)	(0.261,1)	(0.280,1)	(0,2π)	–
	J-divergence	(−532,546)	(−587,561)	(0.113,1)	(0.109,1)	(0,2π)	–
	Improvement	*27.423%*	*33.956%*	*20.027%*	*23.750%*	*0.000%*	*20.449%*
2	MI	(−387,401)	(−419,438)	(0.291,1)	(0.278,1)	(0,2π)	–
	J-divergence	(−452,493)	(−498,504)	(0.166,1)	(0.193,1)	(0,2π)	–
	Improvement	*19.924%*	*16.919%*	*17.630%*	*11.773%*	*0.000%*	*13.132%*
3	MI	(−402,389)	(−383,429)	(0.304,1)	(0.267,1)	(0,2π)	–
	J-divergence	(−521,508)	(−451,535)	(0.172,1)	(0.126,1)	(0,2π)	–
	Improvement	*42.417%*	*30.542%*	*18.966%*	*19.236%*	*0.000%*	*21.403%*
4	MI	(−421,433)	(−395,417)	(0.277,1)	(0.295,1)	(0,2π)	–
	J-divergence	(−519,537)	(−504,539)	(0.161,1)	(0.192,1)	(0,2π)	–
	Improvement	*23.653%*	*28.448%*	*16.044%*	*14.610%*	*0.000%*	*16.133%*
5	MI	(−369,391)	(−388,379)	(0.314,1)	(0.270,1)	(0,2π)	–
	J-divergence	(−449,512)	(−491,457)	(0.228,1)	(0.194,1)	(0,2π)	–
	Improvement	*26.447%*	*23.598%*	*12.536%*	*10.411%*	*0.000%*	*15.808%*

Translation Y, *Scaling X*, *Scaling Y*, and *Rotation*). Meanwhile, J-divergence and mutual information of the deformed image pairs were estimated. The deformations continued until the estimated similarity value of the deformed image pair exceeded the optimum one. This deformation was then considered as the maximum search range (feasible search space) of the corresponding component. Only within this search range can the correct registration be found based on the maximum similarity value. Similarity measure that can provide larger feasible search space is considered to be superior to others.

In three tables, the feasible search space provided by mutual information and J-divergence in five components are presented. Besides, the improvements of J-divergence over mutual information in each component are estimated according to the length the search range. For example, as to the Image Pair #1 in Table 3, the interval of the feasible search space in *Translation X* component is (−272, 266) and (−362, 377), as respectively provided by mutual information and J-divergence. As to mutual information, when the *translation X* component varies within (−272, 266), we can guarantee that mutual information obtains its maximum at 0-displacement (the optimum parameter). The same situation is true with J-divergence within interval (−362, 377). The interval (−362, 377) is larger than (−272, 266) by 37.361%, we consider this as the improvement of J-divergence over mutual information in the *Translation X* component. In addition, we consider the geometric

average of improvements in all components to be overall improvement of the feasible search space, using the Eq. (25) bellow:

$$imp = \left(\prod_{i=1}^{N}(imp_i + 1) \right)^{1/N} \times 100\% - 1 \qquad (25)$$

The imp_i is the improvement in transformation component i, and N is the total number of components of the transformation model, and $N = 5$ in affine transformation model. The overall improvements of J-divergence over mutual information are present in the last column of the tables. For example, as to the SPOT-PalSAR image pair #1 in Table 3, the overall improvement of J-divergence over mutual information in providing feasible search space is 30.492%. According to the compassions of feasible search spaces of 15 selected multimodal image pairs, the overall improvements of J-divergence over mutual information is significant, ranging from 23% to 30% for SPOT-PalSAR image pairs, ranging from 17% 24% for SPOT-TM image pairs, and ranging from 13% to 20% for SPOT-DEM data pairs. Registration models based on J-divergence allow for searching the transformation parameters in a much larger space while guaranteeing the validity of the registration result. In practical application, the superiority of the J-divergence on providing larger feasible search space enables the model to register multimodal image pairs which have smaller scene overlap regions. In other words, registra-

tion models based on J-divergence are much more robust and applicable than models based on mutual information.

5. Conclusions

Entropy-based (e.g., mutual information, known as KL-divergence) image registration models are usually unreliable in practical multimodal image registration tasks, particularly regarding to image pairs with insufficient scene overlap region or greatly misaligned. The similarity measures will become unstable and fail to denote the optimum registration. In this paper, J-divergence, the symmetric form of KL-divergence, was introduced as the similarity measure for multimodal image registration. We mathematically demonstrated that J-divergence is capable of providing a larger feasible search space than mutual information, laying the theoretical foundation that J-divergence is superior for registering image pairs with insufficient scene overlap region.

The J-divergence based registration model was applied to register multimodal images, using a Landsat TM image, a PalSAR image and DEM data as floating images and a SPOT image as the reference one. The performances were compared with registration models based on mutual information and other entropy-based measures such as χ^2-divergence, Lin K-divergence and Kolmogorov distance. The results show that J-divergence is less sensitive to scene overlap ratio than other measures. During the registrations, J-divergence is capable of obtaining a higher value than mutual information if the two images are perfectly aligned and a lower value when there is small overlap region between the image pair. According to the quantitative results, J-divergence can provide a much larger feasible search space than mutual information (13–30% larger), which allows the search algorithm to determine the optimum transformation parameters. The improvement in obtaining larger feasible search space is equivalent to the ability to register greater misaligned or smaller scene overlap image pairs. Thus, the proposed registration model based on J-divergence is more applicable and capable of registering multimodal image pairs with insufficient overlap region.

With the J-divergence as the similarity measure, the model can guarantee that the maximum of the similarity measure will be obtained only when the floating image perfectly aligns, even though the scene overlap of the image pair is not sufficient. The optimization algorithm can search the parameters in much broader ranges without getting tripped in small overlapping situations. However, more efficient optimization algorithms are required when the feasible search space becomes much larger. Moreover, the estimation of image joint histogram during the optimization process is quite time-consuming. Thus, the registration model based on J-divergence is more suitable to register small size images until better approaches are developed for the estimation of image joint histogram. More effort is required towards this way in order to make the method truly operational in practical multimodal image registration tasks.

Acknowledgments

References

Ali, S.M., Silvey, S.D., 1966. A general class of coefficients of divergence of one distribution from another. J. Roy. Stat. Soc.: Ser. B (Methodol.) 28 (1), 131–142.

Almeida-Filho, R., Rosenqvist, A., Shimabukuro, Y.E., Dos Santos, J.R., 2005. Evaluation and perspectives of using multitemporal L-band SAR data to monitor deforestation in the Brazilian Amazˌnia. IEEE Geosci. Remote Sens. Lett. 2 (4), 409–412.

Ardizzone, E., Gallea, R., Gambino, O., Pirrone, R., 2009. Effective and efficient interpolation for mutual information based multimodality elastic image registration. In: IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops), 2009. IEEE, pp. 376–381.

Bach, H., Mauser, W., 2003. Methods and examples for remote sensing data assimilation in land surface process modeling. IEEE Trans. Geosci. Remote Sens. 41 (7), 1629–1637.

Chen, H., Varshnev, P.K., 2004. Size-dependent Image Resampling for Mutual Information Based Remote Sensing Image Registration, pp. 24052408.

Chen, H., Varshney, P.K., 2003. Mutual information-based CT-MR brain image registration using generalized partial volume joint histogram estimation. IEEE Trans. Med. Imaging 22 (9), 1111–1119.

Chen, H., Varshney, P.K., Arora, M.K., 2003a. Performance of mutual information similarity measure for registration of multitemporal remote sensing images. IEEE Trans. Geosci. Remote Sens. 41 (11), 2445–2454.

Chen, H., Arora, M.K., Varshney, P.K., 2003b. Mutual information-based image registration for remote sensing data. Int. J. Remote Sens. 24 (18), 3701–3706.

Chiang, M., et al., 2006. Fluid reglstration of Medical Images Using Jensen-renyi Divergence Reveals 3d Profile of Brain Atrophy in HIV/AIDS, pp. 193196.

Clark, I., Harper, W.V., 2000. Practical Geostatistics. Ecosse North America Columbus, OH.

Cole-Rhodes, A.A., Johnson, K.L., LeMoigne, J., Zavorin, I., 2003. Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient. IEEE Trans. Image Process. 12 (12), 1495–1511.

Collignon, A. et al., 1995. Automated multi-modality image registration based on information theory. Inform. Proc. Med. Imaging 3 (6), 263–274.

CsiszÃr, I., 1963. Eine informationstheoretische Ungleichung und ihre anwendung auf den Beweis der ergodizitðt von Markoffschen Ketten. Publ. Math. Inst. Hungar. Acad. 8, 85–108.

Gao, Z., Gu, B., Lin, J., 2008. Monomodal image registration using mutual information based methods. Image Vis. Comput. 26 (2), 164–173.

Hasan, M., Pickering, M.R., Jia, X., 2012. Robust automatic registration of multimodal satellite images using CCRE with partial volume interpolation. IEEE Trans. Geosci. Remote Sens. 50 (10), 4050–4061.

He, Y., Hamza, A.B., Krim, H., 2003. A generalized divergence measure for robust image registration. IEEE Trans. Signal Process. 51 (5), 1211–1220.

Inglada, J., Giros, A., 2004. On the possibility of automatic multisensor image registration. IEEE Trans. Geosci. Remote Sens. 42 (10), 2104–2120.

Inglada, J., Muron, V., Pichard, D., Feuvrier, T., 2007. Analysis of artifacts in subpixel remote sensing image registration. IEEE Trans. Geosci. Remote Sens. 45 (1), 254–264.

Jeffreys, H., 1939. Theory of Probability. Clarendon Press, Oxford.

Jeffreys, H., 1946. An invariant form for the prior probability in estimation problems. Proc. R. Soc. Lond. A 186 (1007), 453–461.

Jiang, C., Zhang, H., Shen, H., Zhang, L., 2012. A practical compressed sensing-based pan-sharpening method. IEEE Geosci. Remote Sens. Lett. 9 (4), 629–633.

Johnson, K., Cole-Rhodes, A., Zavorin, I., Le Moigne, J., 2001. Mutual information as a similarity measure for remote sensing image registration. Proc. SPIE 4383, Geo-Spatial Image and Data Exploitation II, 51, pp. 51–61.

Karantzalos, K., Sotiras, A., Paragios, N., 2014. Efficient and automated multimodal satellite data registration through MRFs and linear programming. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 329–336.

Kern, J.P., Pattichis, M.S., 2007. Robust multispectral image registration using mutual-information models. IEEE Trans. Geosci. Remote Sens. 45 (5), 1494–1505.

Klein, S., Pluim, J.W., Staring, M., Viergever, M., 2009. Adaptive stochastic gradient descent optimisation for image registration. Int. J. Comput. Vision 81 (3), 227–239.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 22 (1), 79–86.

Launay, M., Guerif, M., 2005. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. Agric. Ecosyst. Environ. 111 (1), 321–339.

Lemieux, L., Jagoe, R., Fish, D.R., Kitchen, N.D., Thomas, D.G., 1994. A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs. Med. Phys. 21 (11), 1749–1760.

Li, J., Cong, R., Jin, L., Wei, P., 2008. A Medical Image Registration Method Based on Weighted Mutual Information, pp. 25492552.

Liang, J., Liu, X., Huang, K., Li, X., Wang, D., Wang, X., 2014. Automatic registration of multisensor images using an integrated spatial and mutual information (SMI) metric. IEEE Trans. Geosci. Remote Sens. 52 (1), 603–615.

Lin, S., Zhu, X., Yan, P., 2009. A novel registration method that incorporates template matching and mutual information. In: IEEE International Conference on Automation and Logistics, 2009. ICAL09. IEEE, pp. 813–818.

Liu, C., Hu, S., Gu, J.J., Yang, J., Yu, M., 2007. Brain image registration based on entropy of mutual information matrix. In: Canadian Conference on Electrical and Computer Engineering. CCECE 2007. IEEE, pp. 1163–1166.

Liu, X. et al., 2010. A new landscape index for quantifying urban expansion using multi-temporal remotely sensed data. Landscape Ecol. 25 (5), 671–682.

Liu, X. et al., 2014. Simulating urban growth by integrating landscape expansion index (LEI) and cellular automata. Int. J. Geograph. Inform. Sci. 28 (1), 148–163.

Lu, X., Zhang, S., Su, H., Chen, Y., 2008. Mutual information-based multimodal image registration using a novel joint histogram estimation. Comput. Med. Imaging Graph. 32 (3), 202–209.

Luo, B., Khan, M.M., Bienvenu, T., Chanussot, J., Zhang, L., 2013. Decision-based fusion for pansharpening of remote sensing images. IEEE Geosci. Remote Sens. Lett. 10 (1), 19–23.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximization of mutual information. IEEE Trans. Med. Imaging 16 (2), 187–198.

Martin, S., Durrani, T.S., 2007. A new divergence measure for medical image registration. IEEE Trans. Image Process. 16 (4), 957–966.

Modersitzki, J., 2004. Numerical Methods for Image Registration. Oxford University Press.

Pardo, M., Vajda, I., 2003. On asymptotic properties of information-theoretic divergences. IEEE Trans. Inf. Theory 49 (7), 1860–1867.

Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A., 2000. Image registration by maximization of combined mutual information and gradient information. IEEE Trans. Med. Imaging 19 (8), 809–814.

Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A., 2004. F-Information measures in medical image registration. IEEE Trans. Med. Imaging 23 (12), 1508–1516.

Richards, J.A., Jia, X., 1999. Remote Sensing Digital Image Analysis: An Introduction. Springer.

Seppa, M., 2008. Continuous sampling in mutual-information registration. IEEE Trans. Image Process. 17 (5), 823–826.

Servello, E.L., Kuplich, T.M., Shimabukuro, Y.E., 2010. Tropical land cover change detection with polarimetric SAR data. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2010. IEEE, pp. 1477–1480.

Silva, L., Bellon, O.R.P., Boyer, K.L., 2005. Precision range image registration using a robust surface interpenetration measure and enhanced genetic algorithms. IEEE Trans. Pattern Anal. Mach. Intell. 27 (5), 762–776.

Studholme, C., Hill, D.L.G., Hawkes, D.J., 1995. Multiresolution voxel similarity measures for MR-PET registration. In: Proceedings of the 14th International Conference on Information Processing in Medical Image (IPMI). Kluwer, Dordrecht, The Netherlands, pp. 287–298.

Studholme, C., Hill, D.L.G., Hawkes, D.J., 1997. Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. Med. Phys. 24 (1), 25–35.

Studholme, C., Hill, D.L.G., Hawkes, D.J., 1999. An overlap invariant entropy measure of 3D medical image alignment. Pattern Recogn. 32 (1), 71–86.

ThÔvenaz, P., Unser, M., 2000. Optimization of mutual information for multiresolution image registration. IEEE Trans. Image Process. 9 (12), 2083–2099.

Topsoe, F., 2000. Some inequalities for information divergence and related measures of discrimination. IEEE Trans. Inf. Theory 46 (4), 1602–1609.

Viola, P., Wells III, W.M., 1997. Alignment by maximization of mutual information. Int. J. Comput. Vision 24 (2), 137–154.

Wachowiak, M., SmolÚkovÃ, R., Peters, T., 2003a. Multiresolution biomedical image registration using generalized information measures. Medical Image Computing and Computer-Assisted Intervention, vol. 2879, pp. 846–853.

Wachowiak, M.P., Smolikova, R., Tourassi, G.D., Elmaghraby, A.S., 2003b. Similarity metrics based on nonadditive entropies for 2 D-3 D multimodal biomedical image registration. In: Proceedings of SPIE. Citeseer, pp. 1090–1100.

Xie, H., Pierce, L.E., Ulaby, F.T., 2003. Mutual information based registration of SAR images. In: Proceedings on Geoscience and Remote Sensing Symposium. IGARSS 2003. IEEE, pp. 4028–4031.

Zhang, L., Shen, S., Gong, W., Zhang, H., 2012. Adjustable model-based fusion method for multispectral and panchromatic images. IEEE Trans. Syst. Man Cyber. Part B Cyber. 42 (6), 1693–1704.

Zitova, B., Flusser, J., 2003. Image registration methods: a survey. Image Vis. Comput. 21 (11), 977–1000.