Received July 27, 2018, accepted September 12, 2018, date of publication September 28, 2018, date of current version October 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2872729

ℓm_p: A Novel Similarity Measure for Matching Local Image Descriptors

GUOHUA LV, (Member, IEEE)
College of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China
e-mail: drguohualv@163.com

This work was supported in part by the Natural Science Foundation of Shandong Province, China, under Grant ZR2017LF005 and Grant ZR2016FM14 and in part by the National Natural Science Foundation of China under Grant 61703219.

ABSTRACT m_p-dissimilarity is a recently proposed data-dependence similarity measure. In the literature, how m_p-dissimilarity is generally used for matching local image descriptors has been formalized, and three matching strategies have been proposed by incorporating ℓ_p-norm distance and m_p-dissimilarity. Each of these three matching strategies is essentially a two-round matching process that utilizes ℓ_p-norm distance and m_p-dissimilarity individually. This paper presents two novel similarity measures for matching local image descriptors. The first similarity measure normalizes and weights the similarities that are calculated using ℓ_p-norm distance and m_p-dissimilarity, respectively. The second similarity measure involves a novel calculation that takes into account both spatial distance and data distribution between descriptors. The proposed similarity measures are extensively evaluated on a few image registration benchmark data sets. Experimental results will demonstrate that the proposed similarity measures achieve higher matching accuracy and are able to attain better recall results when registering multi-modal images compared with the existing matching strategies that combine ℓ_p-norm distance and m_p-dissimilarity.

INDEX TERMS Similarity measure, ℓ_p-norm distance, m_p-dissimilarity, local descriptors, accuracy, image registration

I. INTRODUCTION

Calculating the similarity between two vectors [1] is an essential operation in various applications such as data clustering [2], image retrieval [3] and image registration [5]. The aim of using a similarity measure is to find the closest match between a test instance and instances in a dataset [5].

One of our main research interests is feature based image registration [4]–[8]. In this research field, ℓ_p-norm distance ($p = 2$), i.e. Euclidean distance, has been extensively used [4]–[8], [11]–[16] as it intuitively corresponds to the distance defined in the real three-dimensional world [22]. A data dependency similarity measure called m_p-dissimilarity has been recently proposed in [20]–[22], which was inspired by a distance-density model of dissimilarity measure [18]. The ℓ_p-norm distance calculates the spatial distance in each dimension between two vectors, whereas m_p-dissimilarity considers the relative positions of the two vectors to the rest of data in each dimension. As reported in [20] and [21], m_p-dissimilarity outperforms ℓ_p-norm distance in the applications including data mining tasks including clustering, anomaly detection, and multi-label classification.

In [5], m_p-dissimilarity is formalized and used for matching local image descriptors, and three matching strategies have been proposed by taking into account both ℓ_p-norm distance and m_p-dissimilarity. Essentially, each of these three matching strategies is a two-round matching process which utilizes ℓ_p-norm distance and m_p-dissimilarity individually in matching local image descriptors.

Inspired by the matching strategies proposed in [5], this work is focused on exploring novel similarity measures on the basis of ℓ_p-norm distance and m_p-dissimilarity. Main contributions of this paper are twofold as follows.

1. Proposing a similarity measure by weighting the similarities calculated using ℓ_p-norm distance and m_p-dissimilarity respectively (Section III-A);
2. Proposing a second similarity measure in a new way of calculation which takes into account both spatial distance and data distribution between local image descriptors (Section III-B).

The rest of this paper is organized as follows. Section II briefly reviews the related work. Section III describes two proposed similarity measures in details, followed by

a performance study in Section IV. The paper is concluded in Section V.

II. RELATED WORK

This section briefly reviews ℓ_p-norm distance, m_p-dissimilarity and previously proposed matching strategies which incorporate ℓ_p distance and m_p-dissimilarity.

A. ℓ_p-NORM DISTANCE

Given two vectors $\mathbf{x}$ and $\mathbf{y}$, their ℓ_p-norm distance [17] is defined as

$$\ell_p(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}|| = \left(\sum_{i=1}^{d} |x_i - y_i|^p \right)^{\frac{1}{p}}, \tag{1}$$

where x_i and y_i are the i^{th} components of $\mathbf{x}$ and $\mathbf{y}$, d is the number of dimensions of $\mathbf{x}$ and $\mathbf{y}$, and $||\cdot||_p (p > 0)$ denotes the p order norm of a vector. Eq. 1 represents the well-known Euclidean distance when $p = 2$. The Euclidean distance has been widely used in various applications [4], [30], [31].

B. m_p-DISSIMILARITY

On the basis of mass estimation [19], m_p-dissimilarity has been proposed in [20]–[22] and defined as

$$m_p(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^{d} P_i(\mathbf{z} \in R|\phi(\mathbf{x}))^p \right)^{\frac{1}{p}}, \tag{2}$$

where R denotes a region enclosing $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{z}$ is a point that is randomly selected from R, $\phi(\mathbf{x})$ represents the distribution of data $\mathbf{x}$, $P_i(\mathbf{z} \in R|\phi(\mathbf{x}))$ means the probability that $\mathbf{z}$ falls in R at the i^{th} dimension, and p is a parameter which controls the influence of each dimension by scaling up and down the degree of dissimilarity. Herein, the role of p is similar to that in ℓ_p-norm distance Eq. 1 defines. In [20]–[22], R is a region that is centered at $h = \langle h_1, h_2, \cdots, h_d, \rangle$, where $h_i = \frac{x_i + y_i}{2}$. In Eq. 2, the smaller $m_p(\mathbf{x}, \mathbf{y})$ is, the more similar $\mathbf{x}$ and $\mathbf{y}$ are.

If $\mathbf{x}$ and $\mathbf{y}$ are similar, $m_p(\mathbf{x}, \mathbf{y})$ produces many small $P_i(\mathbf{z} \in R|\phi(\mathbf{x}))$. In practice, as suggested in [20]–[22], $P_i(\mathbf{z} \in R|\phi(\mathbf{x}))$ can be estimated by

$$P_i(\mathbf{z} \in R|\phi(\mathbf{z})) = \frac{|R_i|}{n}, \tag{3}$$

where R_i denotes the region that encloses x_i and y_i, and n is the number of instances in the data. Practically, R_i is set to $[min(x_i, y_i) - \delta, max(x_i, y_i) + \delta]$, where $\delta \geq 0$. With Eqs. 2 and 3, the m_p-dissimilarity of $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$m_p(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^{d} \left(\frac{|R_i|}{n} \right)^p \right)^{\frac{1}{p}}. \tag{4}$$

Similar to ℓ_2 in ℓ_p-norm distance, when $p = 2$, the m_p-dissimilarity is called m_2 for the referencing purpose.

C. MATCHING STRATEGIES INCORPORATING ℓ_p AND m_p

In [5], three matching strategies were proposed by incorporating ℓ_p-norm distance and m_p-dissimilarity in matching local descriptors as follows.

To clearly introduce the matching strategies incorporating ℓ_p and m_p, the following denotations are first given.

i. d: the number of elements in each local descriptor;
ii. N_t: the number of local descriptors built in target image;
iii. N_{mq}: the number of local descriptors in query image which correspond to Euclidean distance based keypoint matches;
iv. N_{mt}: the number of local descriptors in target image which correspond to Euclidean distance based keypoint matches.

Strategy 1: The descriptors in query image are selected and only those descriptors which correspond to Euclidean distance based keypoint matches are used in m_p-dissimilarity based matching. The data for calculating m_p-dissimilarity

$$\mathcal{D}_1 = \mathcal{D}_{mq} \bigcup \mathcal{D}_t, \tag{5}$$

where $\mathcal{D}_{mq}$ denotes those descriptors in query image which correspond to Euclidean distance based keypoint matches, and $\mathcal{D}_t$ represents all descriptors in target image. The dimensionality of $\mathcal{D}_1$ is $d \times (N_{mq} + N_t)$.

Strategy 2: This strategy only uses those descriptors in two images which correspond to Euclidean distance based keypoint matches as the input of calculating m_p-dissimilarity, i.e.

$$\mathcal{D}_2 = \mathcal{D}_{mq} \bigcup \mathcal{D}_{mt}, \tag{6}$$

where $\mathcal{D}_{mq}$ and $\mathcal{D}_{mt}$ are descriptors that correspond to Euclidean distance based keypoint matches in query image and target image, respectively. The dimensionality of $\mathcal{D}_1$ is $d \times (N_{mq} + N_{mt})$.

Strategy 3: First, all descriptors in two images are matched using Euclidean distance to obtain keypoint matches M_{ed}. Second, m_p-dissimilarity is utilized to match all descriptors to attain keypoint matches M_{mp}. Third, these two sets of keypoint matches are intersected to obtain the final keypoint matches M_f, i.e.

$$M_f = M_{ed} \bigcap M_{mp}. \tag{7}$$

Essentially, the aforementioned three matching strategies are a two-round matching process which utilizes ℓ_p-norm distance and m_p-dissimilarity individually. As analyzed in [5], each of these three matching strategies is likely to achieve higher matching accuracy as compared to employing Euclidean distance or m_p-dissimilarity individually. When $p = 2$, these three matching strategies are called ℓm_2^1, ℓm_2^2 and ℓm_2^3, respectively.

III. PROPOSED SIMILARITY MEASURES FOR MATCHING LOCAL DESCRIPTORS

In [5], three matching strategies were proposed by incorporating ℓ_p-norm distance and m_p-dissimilarity. Each of these

three matching strategies is essentially a two-round matching process which utilizes ℓ_p-norm distance and m_p-dissimilarity individually in matching local image descriptors. Different from these matching strategies, this section presents two novel similarity measures. The first proposed similarity measure normalizes ℓ_p-norm distance and m_p-dissimilarity, and then weights these two similarity measures. The calculations for ℓ_p-norm distance and m_p-dissimilarity are simply based on their definitions in Sections II-A and II-B. In contrast, the second proposed similarity measure takes into account spatial distance and data distribution between vectors in a new way of calculation.

A. FIRST SIMILARITY MEASURE

As stated in Sections II-A and II-B, the ℓ_p-norm distance and m_p-dissimilarity can be calculated by Eq. 1 and Eq. 4, respectively. To integrate ℓ_p-norm distance and m_p-dissimilarity into one similarity measure, two operations are performed as follows. The first operation is to normalize the similarity calculated by ℓ_p-norm distance and m_p-dissimilarity, in that these two similarity measures lead to results which are at largely different scales. By doing normalization, the similarity values are all restricted into [0,1], therefore arithmetical operations can be performed on these two similarity measures to potentially give rise to a new similarity measure. Herein, it is noted that a distance calculated by neither ℓ_p-norm distance nor m_p-dissimilarity is negative, therefore it is impossible that any negative value falls into the normalized value range. The second operation is to weight ℓ_p-norm distance and m_p-dissimilarity. Theoretically, it is unknown which of these two similarity measures contributes more to calculating the similarity of vectors accurately. Hence, when defining a new similarity measure, it is essential to give weights on ℓ_p-norm distance and m_p-dissimilarity. Taking into account the two operations analyzed above, the newly-defined similarity measure is as follows:

$$\ell m_p^a(\mathbf{x},\mathbf{y}) = \lambda_1 \frac{\ell_p(\mathbf{x},\mathbf{y})}{\max \ell_p} + \lambda_2 \frac{m_p(\mathbf{x},\mathbf{y})}{\max m_p} = \lambda_1 \frac{\left(\sum_{i=1}^{d} |x_i - y_i|^p\right)^{\frac{1}{p}}}{\max \ell_p} + \lambda_2 \frac{\left(\sum_{i=1}^{d} \left(\frac{|R_i|}{n}\right)^p\right)^{\frac{1}{p}}}{\max m_p}, \tag{8}$$

where $\max \ell_p$ and $\max m_p$ denote the maximum ℓ_p-norm distance and m_p-dissimilarity respectively calculated for all combinations of two vectors in the entire data, and λ_1 and λ_2 are weighting factors for these two similarity measures. The calculation for $\ell_p(\mathbf{x},\mathbf{y})$ and $m_p(\mathbf{x},\mathbf{y})$ can be referred to Eq. 1 and Eq. 4, respectively. For the referencing purpose, the aforementioned similarity measure is called ℓm_p^A. When $p = 2$, it is ℓm_2^A.

B. SECOND SIMILARITY MEASURE

As introduced in Sections II-A and II-B, ℓ_p-norm distance calculates spatial distance between two vectors, whereas m_p-dissimilarity is dependent on how the data is distributed at each dimension of these two compared vectors. Due to the complementarity, the effectiveness of measuring the similarity between vectors is likely to be enhanced by taking into account both spatial distance between compared vectors and the distribution of the entire data. Eq. 9 gives a similarity measure in a new way of calculation as follows:

$$\ell m_p^b(\mathbf{x},\mathbf{y}) = \left(\sum_{i=1}^{d}\left(\lambda_1|x_i - y_i| + \lambda_2\sqrt{\frac{\sum_1^n |z_i - \hat{\mu}|^2}{n}}\right)^p\right)^{\frac{1}{p}}, \tag{9}$$

where x_i and y_i are components at the i^{th} dimension of two vectors being compared, z_i denotes components at the i^{th} dimension across the entire data, $\hat{\mu}$ is the mean value of x_i and y_i, and λ_1 and λ_2 are two weighting factors. In Eq. 9, the two parts at the left and right of $+$ represent spatial distance between compared vectors and the distribution of the entire data, respectively. Similar to Section III-A, the similarity measure defined in Eq. 9 is called ℓm_p^B. When $p = 2$, it is ℓm_2^B.

IV. PERFORMANCE STUDY

One of our main research interests is image registration based on local features [4]–[8], [12], [14]–[16]. This section will evaluate the proposed similarity measures in registering various kinds of images. The compared techniques include ℓm_2^1, ℓm_2^2, ℓm_2^3, ℓm_2^A and ℓm_2^B, which have been introduced in Sections II-C and III. As it has been reported in [5] that ℓm_2^1, ℓm_2^2 and ℓm_2^3 show advantages over ℓ_p and m_p, therefore these two similarity measures will not be compared with the proposed ℓm_2^A and ℓm_2^B. For the purpose of performance comparisons, GO-SIFT and GO-IS-SIFT (GO: Gradient Occurrences, IS: Improved Symmetric) [7] are used as the benchmark feature-based image registration technique for mono-modal and multi-modal images, respectively. Nearest Neighbor Distance Ratio (NNDR) based matching [11], [24], [25] is used for all the five compared techniques.

A. EVALUATION METRICS

The accuracy of an image registration technique depends largely on the matching accuracy. The higher the matching accuracy is, the more accurate the final registration should be [7]. Hence, the proposed matching strategies and similarity measure are evaluated by

$$accuracy = \frac{number\ of\ correct\ matches\ found}{number\ of\ total\ matches\ found} \times 100\%. \tag{10}$$

Moreover, recall vs 1-precision [11] is used for performance evaluation. The precision is simply equivalent of accuracy defined in Eq. 10. The recall is defined as

$$recall = \frac{number\ of\ correct\ matches\ found}{number\ of\ correspondences} \times 100\%. \tag{11}$$

The recall vs 1-precision curve is generally plotted for a particular image pair [5], [11]. To make statistics on a set of image pairs, the area under the recall vs 1-precision curve [23] will be used.

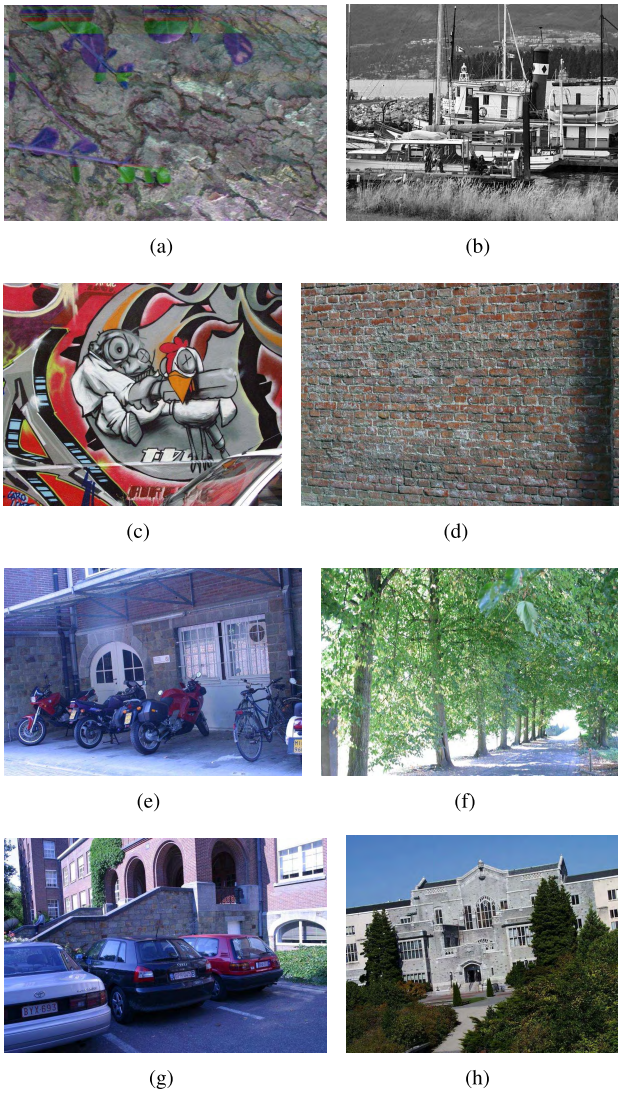


FIGURE 1. Eight base images of the Oxford dataset. (a) bark (scale+rotation). (b) boat (scale+rotation). (c) graffiti (viewpoint) (d) wall (viewpoint). (e) bikes (blur). (f) trees (blur). (g) leuven (illumination). (h) ubc (JPEG compression).

In experiments, the ground-truths of image pairs are all known or provided. A maximum of four pixel error is considered when deciding whether a match is correct or not, which is consistent with existing literature [7], [26].

B. TEST DATASETS

In registering mono-modal images, we use the Oxford dataset [11][1] which is a benchmark dataset in the domain of image registration (Dataset 1). In this dataset, there are five different transformations: scale and rotation, viewpoint, blur, illumination, and JPEG compression. This dataset contains 40 image pairs which stem from eight base images by undergoing an increasing magnitude of transformations. These eight base images are shown in Fig. 1.

[1]The Affine Covariant Regions dataset: http://www.robots.ox.ac.uk/~vgg/data/data-aff.html

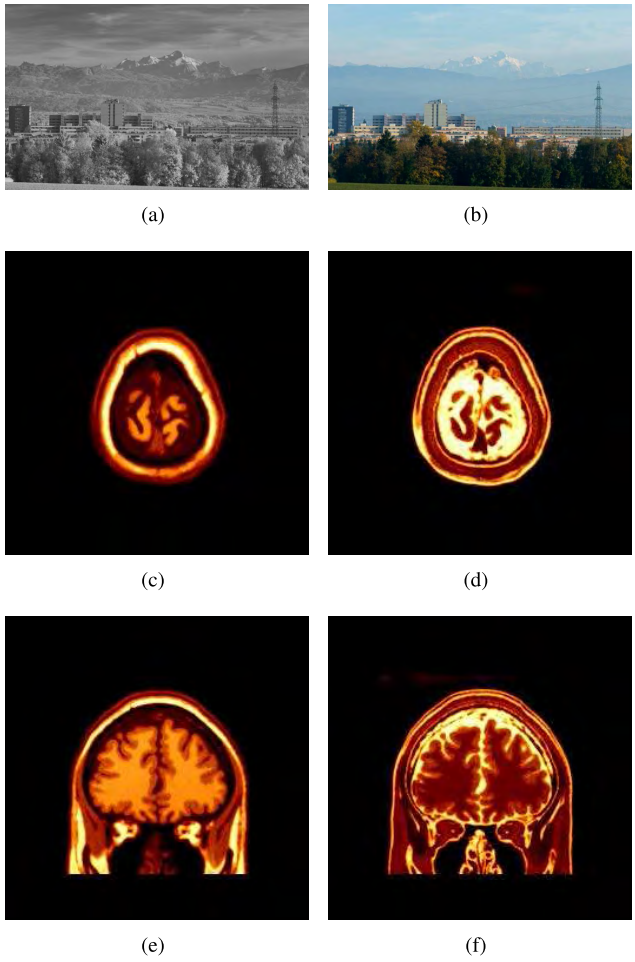


FIGURE 2. Sample image pairs for Datasets 2 to 5. (a) and (b) sample pair for Dataset 2; (c) and (d) sample pair for Dataset 3; (e) and (f) sample pair for Dataset 4.

In registering multi-modal images, the following three datasets are tested. The first dataset consists of 18 NIR (Near Infra-Red) vs EO (Electro-Optical) image pairs from several sources [11], [27]–[29] (Dataset 2). The second and third datasets are transverse and coronal T1 vs T2 weighted MRI brain images, respectively (Datasets 3 and 4). These two datasets were collected from McConnell Brain Imaging Center.[2] There are 87 and 101 image pairs in Datsets 3 and 4, respectively. Fig. 2 shows sample image pairs for Datasets 2 to 4. In total, 246 image pairs are tested in the experiments.

C. COMPARISONS IN ACCURACY

Fig. 3 shows matching accuracy achieved by each of the six compared techniques when registering image pairs of the Oxford dataset. Each sub-figure in Fig. 3 shows matching accuracy for those five pairs associated with the corresponding base image. Fig. 4 (a) to (c) shows matching accuracy when registering multi-modal image pairs of Datasets 2 to 4,

[2]McConnell Brain Imaging Center: https://www.mcgill.ca/bic/home

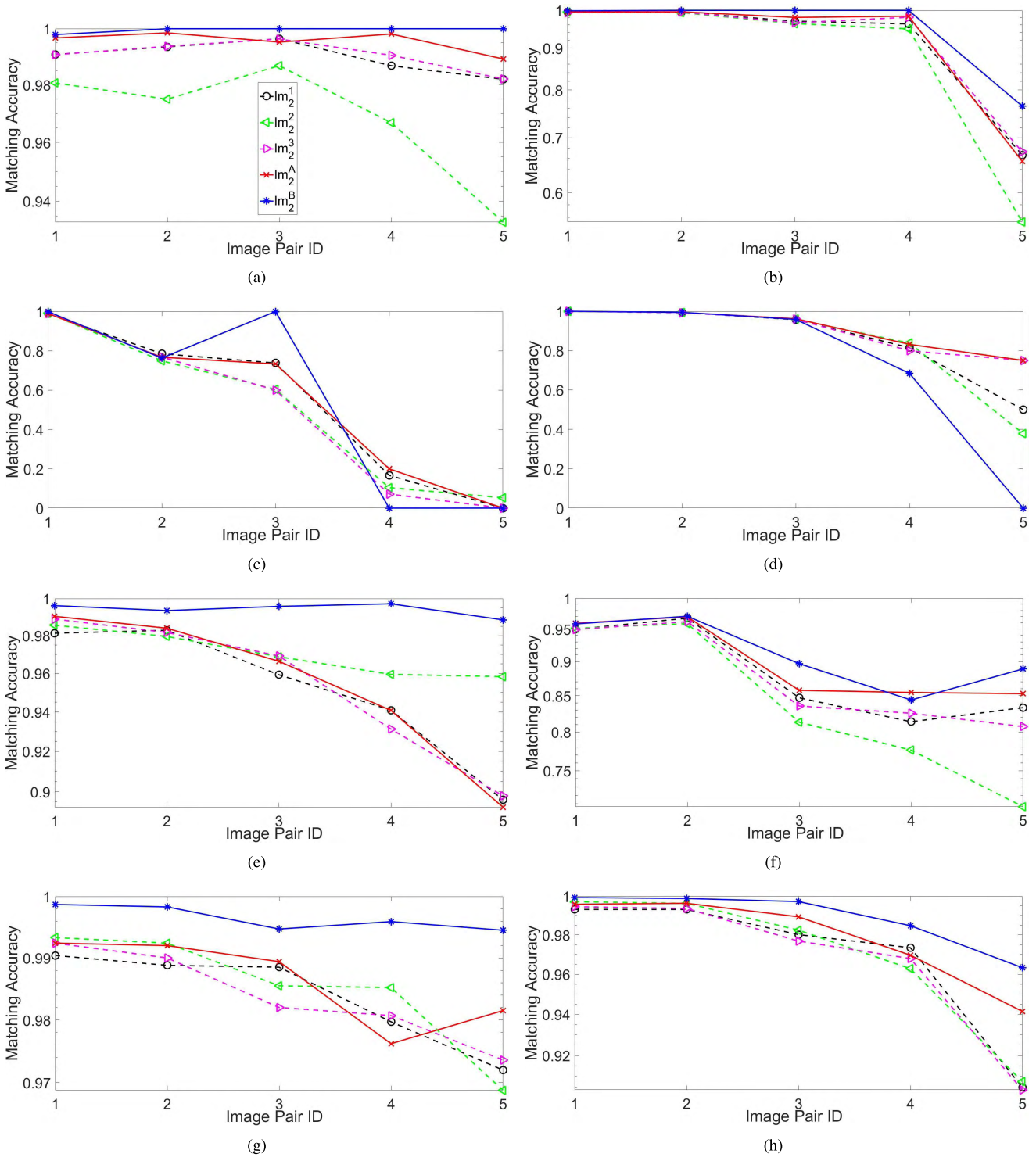
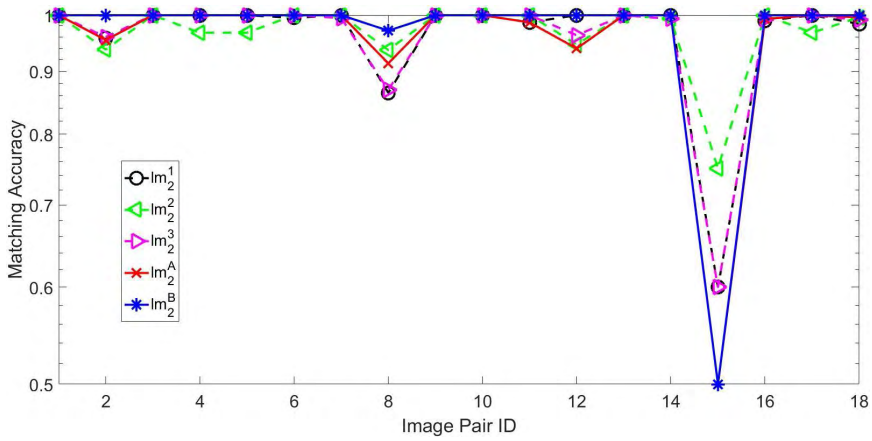


FIGURE 3. Matching accuracy for each pair of the Oxford dataset. (a) bark (scale+rotation). (b) boat (scale+rotation). (c) graffiti (viewpoint) (d) wall (viewpoint). (e) bikes (blur). (f) trees (blur). (g) leuven (illumination). (h) ubc (JPEG compression).

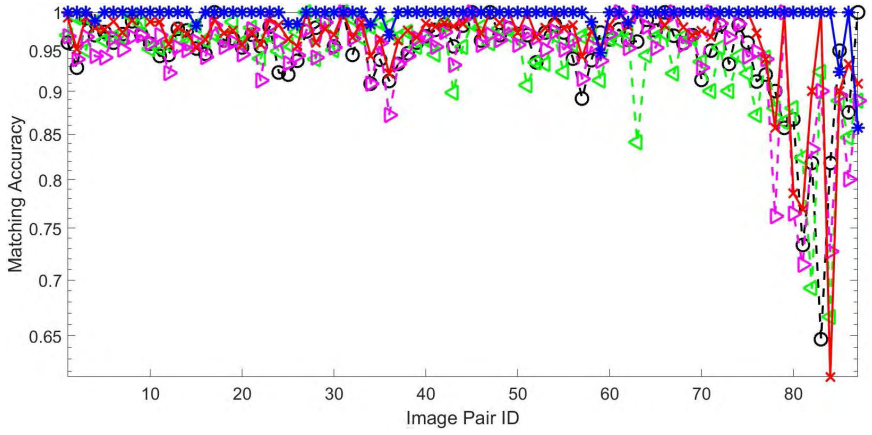
respectively. The average accuracy achieved by each compared technique for each dataset is presented in Table 1.

By observing the results presented in Figs. 3 and 4 as well as Table 1, the following two trends can be drawn.

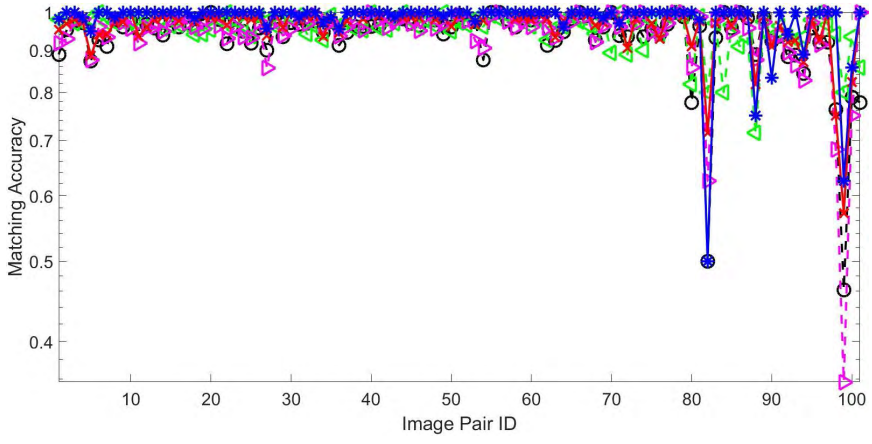
i. On the whole, the proposed ℓm_2^A and ℓm_2^B outperform ℓm_2^1, ℓm_2^2 and ℓm_2^3 in terms of matching accuracy. The average accuracies for all four datasets achieved by ℓm_2^1, ℓm_2^2 and ℓm_2^3 are 93.55%, 93.83% and 93.63%,



(a)



(b)



(c)

FIGURE 4. Accuracy comparisons for Datasets 2 to 4. (a) NIR vs EO (Dataset 2). (b) transverse T1 vs T2 MRI (Dataset 3). (c) coronal T1 vs T2 MRI (Dataset 4).

respectively. By comparison, ℓm_2^A and ℓm_2^B achieve 95.27% and 97.02, respectively.

ii. Overall, ℓm_2^B is able to achieve higher matching accuracy as compared to ℓm_2^A. Across all test image pairs, ℓm_2^B makes a 1.75% accuracy improvement over ℓm_2^A.

Moreover, Fig. 5 compares keypoint matching of the five compared techniques when registering a pair of transverse T1 vs T2 weighted MRI brain images from Dataset 3. With a 94.87% accuracy, ℓm_2^3 performs best among ℓm_2^1, ℓm_2^2 and ℓm_2^3. The accuracy is increased to 97.22% and 100% by the proposed ℓm_2^A and ℓm_2^B, respectively.

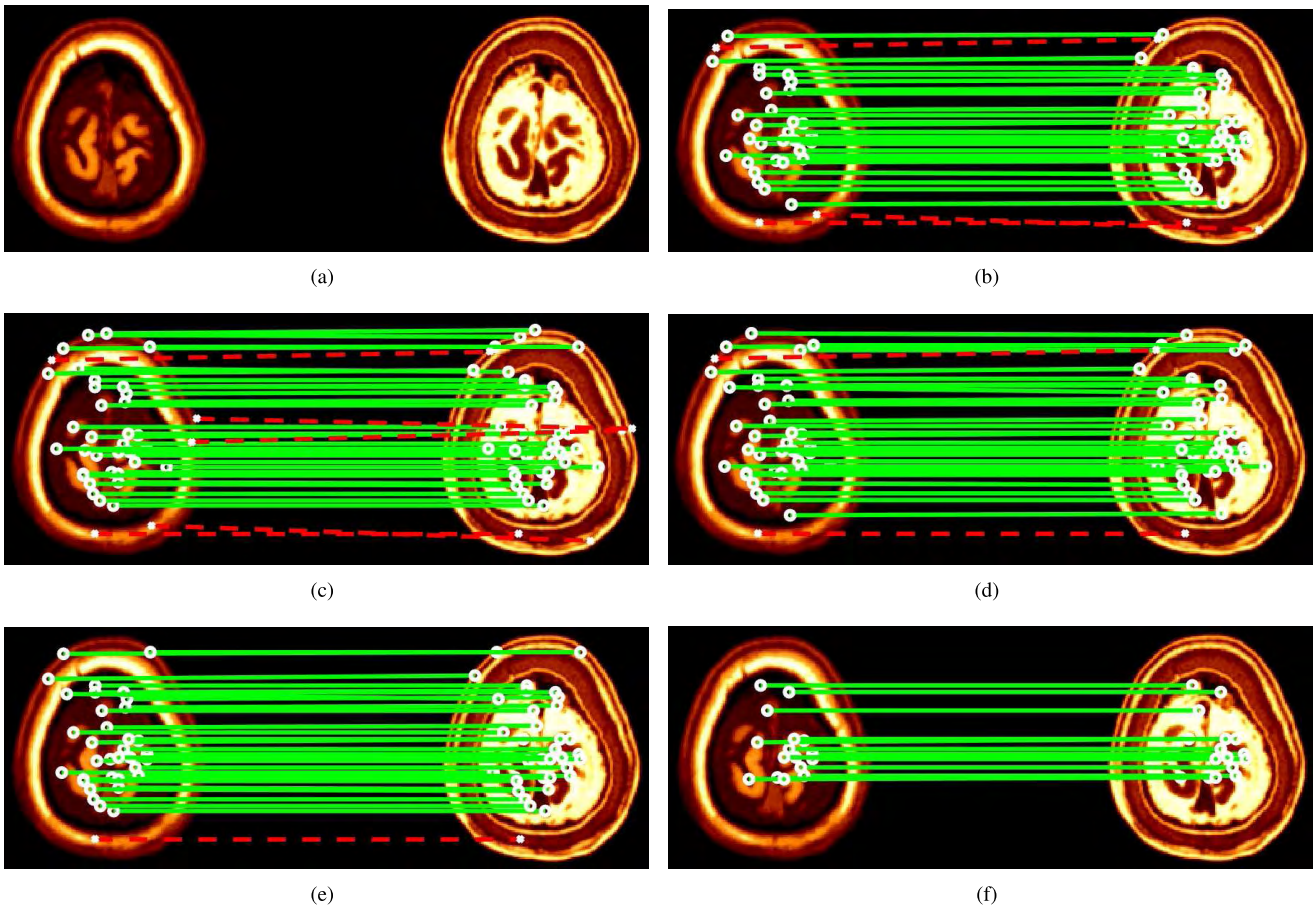


FIGURE 5. Comparisons in matching results for a pair of transverse T1 vs T2 weighted MRI brain images from Dataset 3. Green (solid) and red (dashed) lines indicate correct and incorrect matches respectively. (a) Original image pair. (b) ℓm_2^1: 31/34=91.18%. (c) ℓm_2^2: 34/39=87.18%. (d) ℓm_2^3: 37/39=94.87%. (e) ℓm_2^A: 35/36=97.22%. (f) ℓm_2^B: 17/17=100%. The image pair in (a) is the same as Fig. 2 (b).

TABLE 1. Comparisons in average accuracy of the compared techniques.

Technique	DS1	DS2	DS3	DS4	Ave. Accu.
ℓm_2^1	88.55	96.54	94.81	93.92	93.55
ℓm_2^2	86.99	97.02	94.37	95.50	93.83
ℓm_2^3	88.49	96.47	94.49	94.41	93.63
ℓm_2^A	**89.77**	96.02	96.55	96.20	95.27
ℓm_2^B	89.04	**97.06**	**99.53**	**98.00**	**97.02**

[a] DSi in the first row denotes Dataset i (i=1. . .4).
[b] The accuracy in the last column is obtained by doing average on all four datasets.
[c] For a clear illustration, the percentage symbol, %, is not shown for each accuracy value.
[d] The best result in each column is highlighted in bold.

TABLE 2. Comparisons in area under the recall vs 1-precision curve for the Oxford dataset (Dataset 1). The results achieved by the proposed ℓm_2^A and ℓm_2^B are highlighted in bold. Higher results are better.

Image	ℓm_2^1	ℓm_2^2	ℓm_2^3	ℓm_2^A	ℓm_2^B
bark	0.1912	0.1212	0.2149	**0.4953**	**0.3608**
boat	0.2319	0.1505	0.2503	**0.1780**	**0.1699**
graffiti	0.1464	0.1033	0.1587	**0.0725**	**0.0720**
wall	0.1762	0.1634	0.2049	**0.1084**	**0.1110**
bikes	0.7441	0.5243	0.7485	**0.1776**	**0.1594**
trees	0.2522	0.2399	0.2843	**0.0662**	**0.0637**
leuven	0.7538	0.5797	0.7574	**0.4049**	**0.4414**
ubc	0.3679	0.3543	0.3733	**0.0825**	**0.0735**
average	0.3583	0.2796	0.3740	**0.1982**	**0.1815**

D. COMPARISONS IN RECALL VS 1-PRECISION

Table 2 makes comparisons with regards to the area under the recall vs 1-precision curve [23] for each base image of the Oxford dataset. Each value of the area under the recall vs 1-precision curve in Table 2 is a result summed up over those five image pairs associated with the corresponding base image. The last row of Table 2 shows the averaged results of the compared techniques. Overall, ℓm_2^1, ℓm_2^2 and ℓm_2^3 outperform ℓm_2^A and ℓm_2^B for this dataset.

Fig. 6 compares the recall results of the five compared techniques on Datasets 2 to 4. Note that Fig. 6 (b) only shows the recall results of 18 image pairs which are randomly sampled from Datasets 3 and 4. By showing a part of recall results of Datasets 3 and 4 in Fig. 6 (b), it is clearer for the

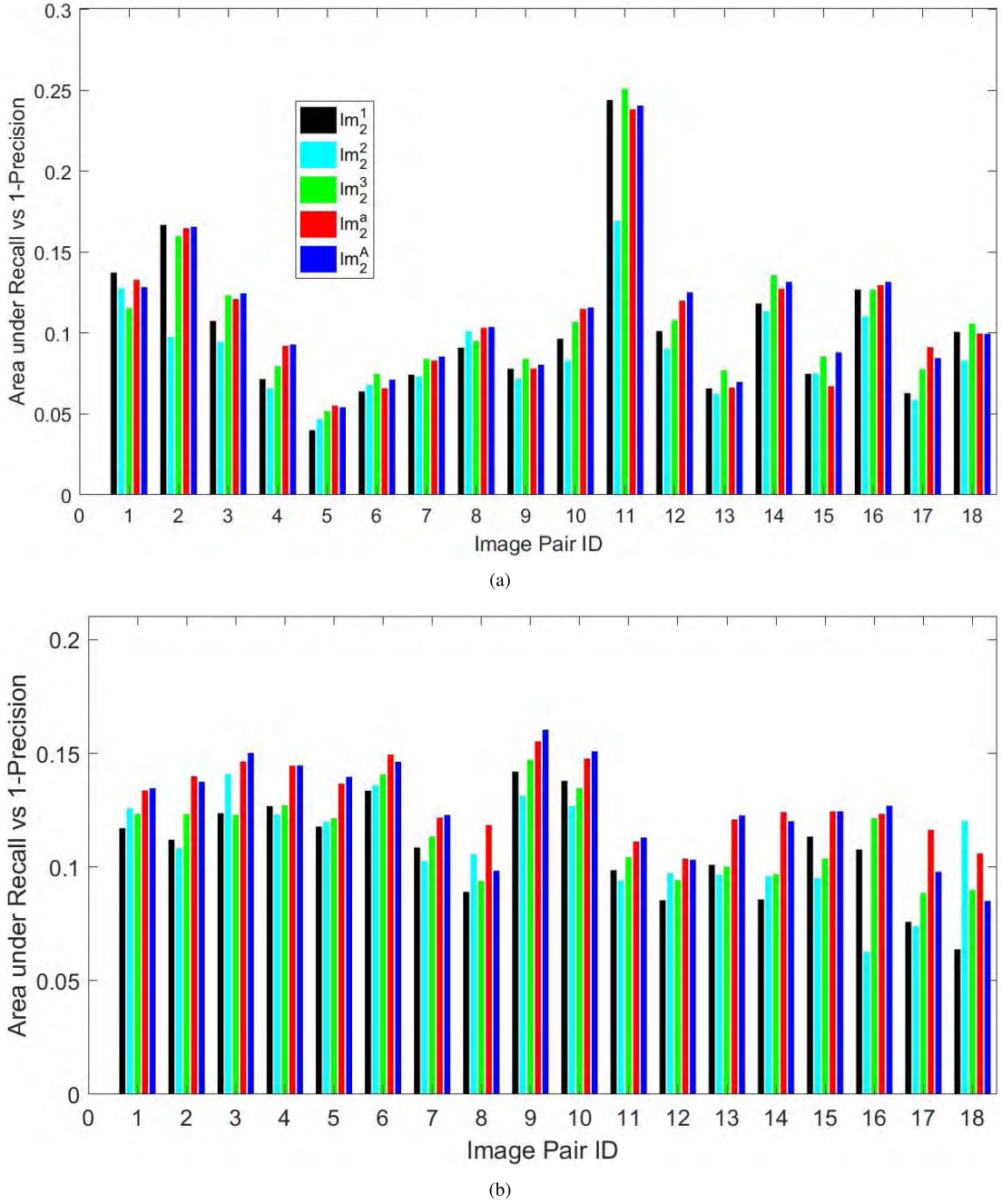


FIGURE 6. Comparisons in area under the recall vs 1-precision curve for Datasets 2 to 4. (a) NIR vs EO (Dataset 2). (b) sample pairs of transverse and coronal T1 vs T2 MRI (Datasets 3 and 4).

illustration purpose. Apart from these sample image pairs, the recall results of the remaining pairs should show a similar trend as Fig. 6 (b), for the reason that all image pairs of Datasets 3 and 4 have similar characteristics. Table 3 lists the averaged area under the recall vs 1-precision curve for Dataset 2 and sampled image pairs of Datasets 3 and 4. As shown in Fig. 6 and Table 3, ℓm_2^A and ℓm_2^B achieve better or comparable recall results as compared to ℓm_2^1, ℓm_2^2 and ℓm_2^3. Moreover, Fig. 7 shows the recall vs 1-precision curves for three sample image pairs from Datasets 2 to 4, which shows a similar performance trend as Table 3.

E. WEIGHTING ISSUE IN PROPOSED SIMILARITY MEASURES

The proposed similarity measures ℓm_2^A and ℓm_2^B have two weighting factors, i.e. λ_1 and λ_2, as stated in Eqs. 8 and 9. In the experimental results shown in Sections IV-C and IV-D, $\lambda_1 = 0.50$ and $\lambda_2 = 0.50$. This is based on the intuition that

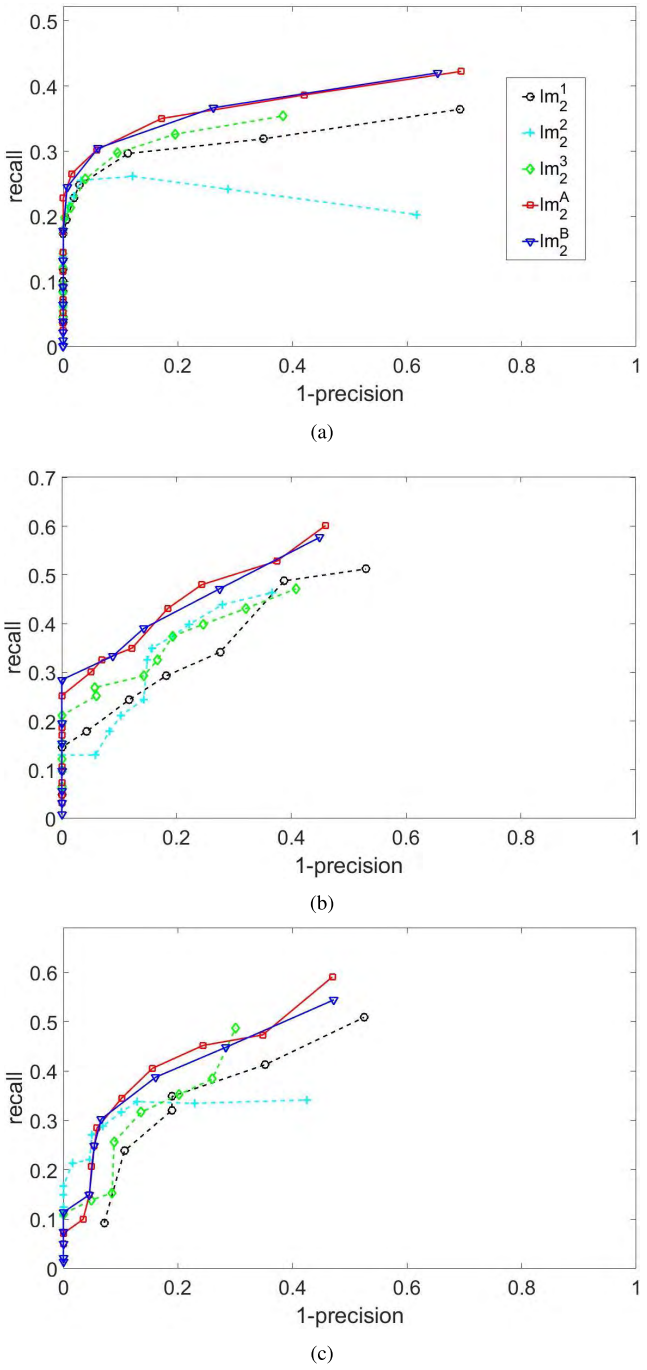


FIGURE 7. Comparisons in recall vs 1-precision. (a) The 3^{rd} pair of NIR vs EO (Dataset 2). (b) The 76^{th} pair of transverse MRI (Dataset 3). (c) The 36^{th} pair of coronal MRI (Dataset 4). (a), (b) and (c) correspond to the sample image pairs shown in Fig. 2 (a), (b) and (c), respectively.

spatial distance and data distribution between vectors are of equal importance in the proposed similarity measures.

To find the optimal choice for these two weighting factors, a set of other λ_1 and λ_2 patterns have been tested, including 0.9 vs 0.1, 0.8 vs 0.2, 0.7 vs 0.3, 0.6 vs 0.4, 0.4 vs 0.6, 0.3 vs 0.7, 0.2 vs 0.8 and 0.1 vs 0.9. Tables 4 and 5 compare the matching accuracy achieved by each λ_1 and λ_2 pattern for ℓm_2^A and ℓm_2^B, respectively.

TABLE 3. Comparisons in area under the recall vs 1-precision curve for Datasets 2 to 4. Randomly sampled image pairs of Datasets 3 and 4 are tested. The results achieved by the proposed ℓm_2^A and ℓm_2^B are highlighted in bold. Higher results are better.

Dataset	ℓm_2^1	ℓm_2^2	ℓm_2^3	ℓm_2^A	ℓm_2^B
2	0.1012	0.0885	0.1079	**0.1084**	**0.1108**
3&4	0.1078	0.1087	0.1137	**0.1291**	**0.1266**

TABLE 4. Accuracy comparisons when various patterns of λ_1 vs λ_2 are used in ℓm_2^A.

λ_1 vs λ_2	DS1	DS2	DS3	DS4	Ave. Accu.	#Correct
0.9 vs 0.1	84.35	94.38	92.45	92.31	91.22	309.63
0.8 vs 0.2	86.47	94.74	94.00	93.73	92.72	297.11
0.7 vs 0.3	88.06	95.44	95.02	95.01	93.92	285.07
0.6 vs 0.4	89.39	95.76	96.17	95.77	94.87	272.92
0.5 vs 0.5	89.77	96.02	96.55	96.20	95.27	261.44
0.4 vs 0.6	91.24	96.41	97.17	96.51	95.88	250.32
0.3 vs 0.7	91.25	96.69	97.94	96.96	96.36	240.24
0.2 vs 0.8	90.85	96.73	98.09	97.11	96.41	230.67
0.1 vs 0.9	90.94	96.81	98.11	97.49	96.59	222.22

[a] DSi in the first row denotes Dataset i (i=1. . .4).
[b] Ave. Accu. in the second last column denotes the average accuracy on all four datasets.
[c] #Correct in the last column represents the average number of correct matches on all four datasets.
[d] For a clear illustration, the percentage symbol, %, is not shown for each accuracy value.

Table 4 shows how ℓm_2^A performs when various λ_1 and λ_2 patterns are used. The second last column of Table 4 lists the matching accuracy averaged on all four datasets. Clearly, the accuracy increases progressively as λ_1 decreases or λ_2 goes up. Meanwhile, the number of correct matches decreases, as shown in the last column of Table 4. Considering both matching accuracy and the number of correct matches, it would be a good choice to use 0.50 for both λ_1 and λ_2.

TABLE 5. Accuracy comparisons when various patterns of λ_1 vs λ_2 are used in ℓm_2^B. The remarks of Table 4 apply in this table.

λ_1 vs λ_2	DS1	DS2	DS3	DS4	Ave. Accu.	#Correct
0.9 vs 0.1	84.38	94.50	92.43	92.52	91.31	308.82
0.8 vs 0.2	87.16	95.03	94.35	94.10	93.13	291.32
0.7 vs 0.3	89.48	95.98	96.25	95.84	94.96	265.76
0.6 vs 0.4	90.84	96.75	98.10	97.01	96.37	226.52
0.5 vs 0.5	89.04	97.06	99.53	98.00	97.02	166.70
0.4 vs 0.6	84.72	90.87	96.55	95.44	93.76	78.76
0.3 vs 0.7	55.00	22.22	37.93	40.59	40.65	4.01
0.2 vs 0.8	0.00	0.00	0.00	0.00	0.00	0.00
0.1 vs 0.9	0.00	0.00	0.00	0.00	0.00	0.00

Likewise, Table 5 compares the matching accuracy and number of correct matches when various λ_1 and λ_2 patterns are used in ℓm_2^B. Interestingly, ℓm_2^B performs best when $\lambda_1 = 0.50$ and $\lambda_2 = 0.50$. When $\lambda_1 < 0.50$ or $\lambda_2 > 0.50$, both matching accuracy and the number of correct matches decline dramatically. Thus, it is a good choice for ℓm_2^B to set both λ_1 and λ_2 to 0.50 as well.

F. COMPARISONS IN EFFICIENCY

Although this work focuses on exploring novel similarity measures, the runtime of the compared techniques has been recorded. The experiments were conducted in Matlab R2014b on a Windows 10 laptop with Intel Core i7 CPU of 2.6GHz and 12GB memory. For all 246 image pairs of Datasets 1 to 4, the average runtime for ℓm_2^1, ℓm_2^2, ℓm_2^3, ℓm_2^A and ℓm_2^B is 81.63, 29.66, 201.65, 100.94 and 99.92 seconds, respectively. Since the experiments were carried out in Matlab, the efficiency should be significantly improved on some other programming platforms such as C and/or C++.

G. DISCUSSIONS

The first proposed similarity measure, i.e. ℓm_2^A, normalizes ℓ_p-norm distance and m_p-dissimilarity, and then weights these two similarity measures. The calculations for ℓ_p-norm distance and m_p-dissimilarity are simply based on their definitions. By comparison, the second proposed similarity measure, i.e. ℓm_2^B, takes into account spatial distance and data distribution between vectors in a new way of calculation.

Compared with the existing matching strategies that combine ℓ_p-norm distance and m_p-dissimilarity, the proposed similarity measures achieve higher matching accuracy in registering various kinds of mono-modal and multi-modal images. With regards to the recall vs 1-precision performance, the proposed similarity measures perform worse in registering mono-modal images, whereas are able to achieve better performance in registering multi-modal images. It is believed that the greater robustness the proposed similarity measures have shown to multi-modal images arise from the normalization operation for ℓm_2^A and a new way of similarity calculation between vectors for ℓm_2^B.

V. CONCLUSIONS

In this paper, two novel similarity measures called ℓm_2^A and ℓm_2^B have been presented for matching local image descriptors. The widely-used ℓ_p-norm distance and recently-proposed m_p-dissimilarity are the foundation of the proposed similarity measures. In the literature, there exist three matching strategies that incorporate ℓ_p-norm distance and m_p-dissimilarity. Each of these three matching strategies is essentially a two-round matching process which utilizes ℓ_p-norm distance and m_p-dissimilarity individually in matching local image descriptors. Inspired by these three matching strategies, this work aims to explore novel similarity measures on the basis of ℓ_p-norm distance and m_p-dissimilarity. A distinct difference from the three existing matching strategies is that local descriptors are only matched once when utilizing the proposed similarity measures.

As the experimental results have shown, the proposed ℓm_2^A and ℓm_2^B are capable of achieving higher matching accuracy as compared to three existing matching strategies that combine ℓ_p-norm distance and m_p-dissimilarity. Moreover, the proposed similarity measures attains better recall vs 1-precision performance when registering multi-modal images.

Without loss of generality, the proposed similarity measures are applicable to the research problems that demand matching local image descriptors. The source code of the proposed technique was written in MATLAB and will be available at https://www.researchgate.net/profile/Guohua_Lv3.

REFERENCES

[1] A. A. Goshtasby, "Similarity and dissimilarity measures," in *Image Registration* (Advances in Computer Vision and Pattern Recognition). London, U.K.: Springer, 2012.
[2] A. S. Shirkhorshidi, S. Aghabozorgi, and T. Y. Wah, "A comparison study on similarity and dissimilarity measures in clustering continuous data," *PLoS ONE*, vol. 10, no. 12, p. e0144059, Dec. 2015.
[3] L. Jiao, X. Tang, B. Hou, and S. Wang, "SAR images retrieval based on semantic classification and region-based similarity measure for earth observation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 3876–3891, Sep. 2015.
[4] G. Lv, "A novel correspondence selection technique for affine rigid image registration," *IEEE Access*, vol. 6, pp. 32023–32034, 2018.
[5] G. Lv, S. W. Teng, and G. Lu, "Enhancing image registration performance by incorporating distribution and spatial distance of local descriptors," *Pattern Recognit. Lett.*, vol. 103, pp. 46–52, Feb. 2018.
[6] G. Lv, S. W. Teng, and G. Lu, "COREG: A corner based registration technique for multimodal images," *Multimedia Tools Appl.*, vol. 77, no. 10, pp. 12607–12634, May 2018.
[7] G. Lv, S. W. Teng, and G. Lu, "Enhancing SIFT-based image registration performance by building and selecting highly discriminating descriptors," *Pattern Recognit. Lett.*, vol. 84, pp. 156–162, Dec. 2016.
[8] G. Lv, "Robust and effective techniques for multi-modal image registration," Ph.D. dissertation, Gippsland School Inf. Technol., Monash Univ., Melbourne, VIC, Australia, 2015.
[9] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, pp. 977–1000, Oct. 2003.
[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[11] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
[12] J. Chen and Tian, "Real-time multi-modal rigid registration based on a novel symmetric-SIFT descriptor," *Prog. Natural Sci.*, vol. 19, no. 5, pp. 643–651, May 2009.
[13] J. Chen, J. Tian, N. Lee, J. Zheng, R. T. Smith, and A. F. Laine, "A partial intensity invariant feature descriptor for multimodal retinal image registration," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1707–1718, Jul. 2010.
[14] S. W. Teng, M. T. Hossain, and G. Lu, "Multimodal image registration technique based on improved local feature descriptors," *J. Electron. Imag.*, vol. 24, no. 1, pp. 013013-1–013013-17, Jan. 2015.
[15] M. T. Hossain, "An effective technique for multi-modal image registration," Ph.D. dissertation, Gippsland School Inf. Technol., Monash Univ., Melbourne, VIC, Australia, 2012.
[16] M. T. Hossain, G. Lv, S. W. Teng, G. Lu, and M. Lackmann, "Improved symmetric-SIFT for multi-modal image registration," in *Proc. DICTA*, Dec. 2011, pp. 197–202.
[17] M. Deza and E. Deza, *Encyclopedia of Distances*. Berlin, Germany: Springer, 2009.
[18] C. L. Krumhansl, "Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density," *Psychol. Rev.*, vol. 85, no. 5, pp. 445–463, Sep. 1978.
[19] C. Bo, K. M. Ting, T. Washio, and G. Haffari, "Half-space mass: A maximally robust and efficient data depth method," *Mach. Learn.*, vol. 100, nos. 2–3, pp. 677–699, Aug. 2015.
[20] S. Aryal, K. M. Ting, T. Washio, and G. Haffari, "Data-dependent dissimilarity measure: An effective alternative to geometric distance measures," *Knowl. Inf. Syst.*, vol. 53, no. 2, pp. 479–506, Apr. 2017.
[21] K. M. Ting, Y. Zhu, M. J. Carman, Y. Zhu, and Z.-H. Zhou, "Overcoming key weaknesses of distance-based neighbourhood methods using a data dependent dissimilarity measure," in *Proc. KDD*, Aug. 2016, pp. 1205–1214.

[22] S. Aryal, K. M. Ting, G. Haffari, and T. Washio, "m_p-dissimilarity: A data dependent dissimilarity measure," in *Proc. ICDM*, Dec. 2014, pp. 707–712.
[23] G. Levi and T. Hassner, "LATCH: Learned arrangements of three patch codes," in *Proc. WACV*, Mar. 2016, pp. 1–9.
[24] B. Fan, F. Wu, and Z. Hu, "Rotationally invariant descriptors using intensity order pooling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 2031–2045, Oct. 2012.
[25] X. Fengguang and H. Xie, "A 3D surface matching method using keypoint-based covariance matrix descriptors," *IEEE Access*, vol. 5, pp. 14204–14220, 2017.
[26] G. H. Yang, C. V. Stewart, M. Sofka, and C.-L. Tsai, "Registration of challenging image pairs: Initialization, estimation, and decision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1973–1989, Nov. 2007.
[27] A. Kelman, M. Sofka, and C. V. Stewart, "Keypoint descriptors for matching across multiple image modalities and non-linear intensity variations," in *Proc. CVPR*, Jun. 2007, pp. 1–7.
[28] Y. S. Kim, J. H. Lee, and J. B. Ra, "Multi-sensor image registration based on intensity and edge orientation information," *Pattern Recognit.*, vol. 41, no. 11, pp. 3356–3365, 2008.
[29] C. Fredembach and S. Susstrunk, "Illuminant estimation and detection using near-infrared," in *Proc. Digit. Photogr. V, IS&T-SPIE Electron. Imag. Symp.*, Jan. 2009.
[30] Z. Yu, W. Chen, X. Guo, X. Chen, and C. Sun, "Analog network-coded modulation with maximum Euclidean distance: Mapping criterion and constellation design," *IEEE Access*, vol. 5, pp. 18271–18286, 2017.
[31] A. Ananth and M. D. Selvaraj, "Error analysis of SSK with Euclidean distance based selection combining," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3195–3204, Apr. 2018.

GUOHUA LV received the Ph.D. degree from Monash University, Australia, in 2015. He is currently with the College of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China. As the first author, he has published six articles in international journals including the IEEE Access, the *Pattern Recognition Letters*, and the *Multimedia Tools and Applications*. His research interests are broadly in image processing, especially in feature-based image registration and hyperspectral image processing. He is a member of the IEEE Computer Society. He is also a Reviewer for some international journals, such as the IEEE Access and the *IET Image Processing*.