Pattern Recognition Letters 128 (2019) 536–543

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Combining dissimilarity measures for image classification

Chuanyi Liu, Junqian Wang, Shaoming Duan, Yong Xu*

Bio-Computing Research Center, College of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), 518055, Shenzhen, China

ARTICLE INFO

Article history:
Received 19 August 2019
Revised 9 October 2019
Accepted 22 October 2019
Available online 25 October 2019

Keywords:
Image classification
Distance measures
Fusion distance
Pattern recognition

ABSTRACT

The local dissimilarity has been verified as one of effective metrics for pattern classification. For high-dimensional data, because of the property of high-dimension, even if there are a number of available samples, they are still only resultant observations of a sampling process of the high-dimensional population. As a consequence, available samples at least partially possess the property of randomness and are not "accurate" representations of the true and total sample space. Besides the prevailing local dissimilarity measure, global dissimilarity measures might also be exploited for improving the classification approach. In this paper, we propose to directly exploit global and local dissimilarity measures to efficiently perform image classification. The proposed method proposes to simultaneously use three dissimilarities derived from the original and transform sample space. These dissimilarities including the elaborated distance ratio enable space relations of the probe sample and gallery samples to be measured from three viewpoints, so the combination of them provide us with more reliable measurements on spatial geometric relationship of samples. An obvious advantage of this combination is that we can attain a very robust evaluation on the space distance between samples and the consequent classification decision will be less affected by the noise in data. The experiments prove that the proposed method does achieve the desired goal, i.e., very satisfactory accuracy improvement in comparison with the previous state-of-the art methods.

1. Introduction

Classification of images is not only one of the most important computer vision problems but also an important means to make our society intelligent [2,5,16,20]. To implement classification of images, dissimilarity measures must be used. A reasonable idea is that a probe sample can be viewed as a member of the class of the gallery with which the probe sample has the lowest dissimilarity [4,9]. Nearest neighbor classification (NNC) is a well-known specific implementation of this idea [3,27].

Extensive applications of NNC are mainly attributed to not only its simplicity but also its following theoretical property. Under the condition that there are infinite gallery samples, the rate of classification errors of NNC must be lower than twice of that of the minimum-error Bayesian classification method (MEBCM) [7]. Because MEBCM is the optimal classifier with the fewest errors, NNC of course is a popular method owing to its good theoretical performance. However, in most real-world applications infinite samples even a large number of samples are unavailable. Especially, for applications on image classification, it seems that in most cases the number of the gallery samples are less than the dimensionality [21]. In this sense, NNC usually cannot exert its ideal performance owing to a limited number of gallery samples.

The center distance seems to be a feasible means to denote global dissimilarity of a probe sample and a class. When the center distance is used for classification, the computational cost is very low. Moreover, it appears that the center distance based classification method might be the fastest classification method, because only calculation of C distance values is needed for classifying a probe sample (C is the number of classes). If all classes are separated from each other, then it is undoubted that the centers of different classes are far from each other. As a result, the center distance based classification method can produce a high accuracy. However, we must point out that in real world applications different classes may have complex distributions. For example, distributions of some classes may not be subject to the normal distribution and even are multimodal [19]. Under these circumstances, the center distance based classification method usually cannot obtain a satisfactory accuracy.

In past studies, people pay a lot of attentions on choosing or designing suitable dissimilarity metrics to represent the distance. For example, we see that metric learning is a popular topic [31,33]. In recent years algorithm based dissimilarity metrics also receive many concerns. The sparse representation based dissimilarity metric is a typical example of algorithm based dissimilarity metrics [25,29]. In the corresponding method, sparse representation is first

* Corresponding author.
E-mail address: laterfall@hit.edu.cn (Y. Xu).

https://doi.org/10.1016/j.patrec.2019.10.026
0167-8655/

implemented, then the solution of the sparse representation is exploited to calculate dissimilarities of the probe sample with every class. Besides the original sparse representation method, the l_2 norm regularization based sparse representation methods including collaborative representation also use the same or similar dissimilarity metric [6,14,26,28,30]. However, people almost ignore a fact that different dissimilarity metrics all contain useful information on representations of space relations of samples. As a result, in this paper, we propose to simply combine three different dissimilarity metrics for efficient image classification.

As we know, most of methods extract features from original samples before pattern classification [32]. The main goal of feature extraction is to attain features beneficial to classification. For example, subspace learning methods including linear discriminant analysis and principal component analysis are widely used feature extraction methods [10,12,18,24]. Subspace learning also has the goal to reduce the dimension of the raw data. As a kind of typical feature transform method, subspace learning converts raw data into a new space [13]. Besides feature transform, feature selection is also widely used [22]. Feature selection can perform well under the condition that some components of the original features are better in representing discriminative information of patterns than the other components [23]. However, with this paper, we find that in the original sample space, if we choose proper dissimilarity metrics, we can still attain very good classification result. Therefore, we propose an image classification method without feature extraction. In particular, though our proposed method copes with the classification problem of images by virtue of only the raw data, it can attain very high classification accuracies. In this sense, for image classification, conventional feature extraction may be not a necessary procedure. This can be partially attributed to the fact that all information obtained using feature extraction methods indeed exists in the raw data.

Our method also has the following implications. Different dissimilarity metrics have different physical meaning. Moreover, their performance vary with the applications. Differing from metric learning methods, our method demonstrates that combination of multiple dissimilarity metrics is very useful for classification of images. Additionally, our findings also show that if the combined dissimilarity metrics have enough difference, the result will be good.

2. The proposed method

The basic idea of the proposed method is that both global and local dissimilarity measures of data are useful for representations of space relations of samples so combination of them can improve the classification performance.

Suppose that all images are gray images. The proposed method has the following main calculations. First, all gallery samples and probe samples are transformed into column vectors. So we also call each of them a sample vector. In particular, as each naïve sample is an image, we first concatenate all columns of an image to generate a column vector. Then the Euclidean distance between each original probe sample and original gallery sample is calculated. The Euclidean distance between probe sample t and the jth gallery sample of the ith class is denoted by e_i^j. Then all gallery samples and probe samples are normalized as unit vectors of length 1. In other words, sample vector g is normalized by $h = \frac{g}{||g||}$. The resultant h is called transform sample. Hereafter we call the sample before the normalization procedure original sample. The space corresponding to the original samples and transform samples are referred to as original sample space and transform sample space respectively. The center of all transform samples of a class is obtained using

$$p_i = \frac{1}{n}\sum_{j=1}^{n} s_i^j \tag{1}$$

p_i denotes the center of the ith class. s_i^j stands for the jth gallery sample of the ith class. n is the number of the gallery samples of the ith class. The center distance between probe sample t and the ith class is defined as

$$c_i = \sqrt{\sum_{k=1}^{M}(p_{ik} - t_k)^2} \tag{2}$$

p_{ik} is the kth entry of p_i. t_k is the kth entry of t. M is the dimension of the sample. The block distance between probe sample t and the jth gallery sample of the ith class is denoted by f_i^j and defined as

$$f_i^j = \sum_{k=1}^{M} |s_{ik}^j - t_k| \tag{3}$$

s_{ik}^j is the kth entry of s_i^j. Based on the basic calculations presented above, we offer all steps on probe sample t of the proposed method as follows.

Step 1. The Euclidean distance between original probe sample t and the original jth gallery sample of the ith class is obtained and denoted by e_i^j. We define $e_i^{min} = \min_{j=1\cdots n} e_i^j$ as Euclidean distance between original probe sample t and the ith class. e_i^{min} is also called the nearest distance of the probe sample to the ith class in the original sample space. For original probe sample t, the ratio of the Euclidean distance of the ith class to the sum of the Euclidean distances of all other classes is attained using $e_{ratio}^i = \frac{e_i^{min}}{\sum_{j=1}^{C} e_j^{min} - e_i^{min}}$. C is still the number of the classes. e_{ratio}^i is called original distance ratio of the probe sample with respect to the ith class.

Step 2. In the transform sample space, the center distance between probe sample t and the ith class is attained using Eq. (2) and denoted by c_i.

Step 3. In the transform sample space, the block distance between probe sample t and the j-th gallery sample of the ith class is obtained using Eq. (3) and denoted by f_i^j. Let $f_i^{min} = \min_{j=1\cdots n} f_i^j$ stand for the block distance between probe sample t and the ith class. f_i^{min} is also called the nearest block distance of the probe sample to the ith class in the transform sample space.

Step 4. The obtained dissimilarity values are normalized by $e_{ratio}^i = \frac{e_{ratio}^i}{max(e_{ratio}^i)}$, $c_i = \frac{c_i}{max(c_i)}$ and $f_i^{min} = \frac{f_i^{min}}{max(f_i^{min})}$, respectively. The resultant e_{ratio}^i, c_i and $f_i^{\min}$ are called normalized distance ratio, normalized center distance and normalized block distance, respectively. The ultimate dissimilarity between probe sample t and the ith class is attained using $u_i = r_1 \times e_{ratio}^i + r_2 \times c_i + r_3 \times f_i^{\min}$. u_i is called fusion distance. r_1, r_2 and r_3 are the weight coefficients to combine three distances e_{ratio}^i, c_i and $f_i^{\min}$.

Step 5. Let $q = \underset{i}{\operatorname{argmin}}\, u_i$. Then we think that probe sample t belongs to the q-th class and denote the classification decision by $label(t) = q$.

Fig. 1 summarizes the pipeline of the proposed method. It clearly presents that one dissimilarity metric is derived from the original sample space and two dissimilarity metrics are generated from the transform sample space.

3. Analysis of our proposed method

With this section, we present the characteristics and rationale of the proposed method. The first characteristic of the proposed method is that it evaluates dissimilarities between probe samples and gallery samples in both the original sample space and transform sample space. The transform sample space is very easy to obtain from the original sample space, and it converts the components of the original sample vector into a fixed range by the nor-

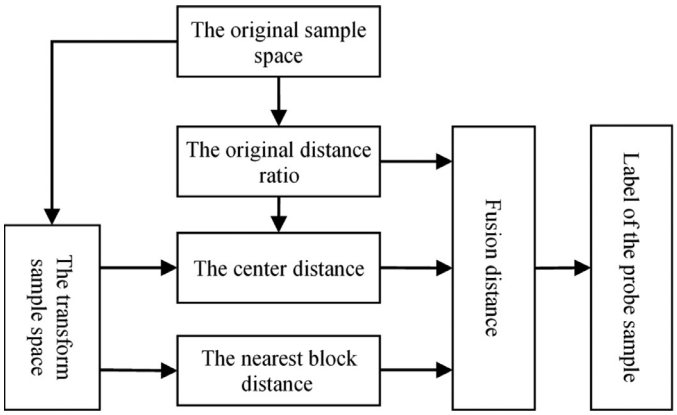

Fig. 1. The pipeline of the proposed method.

malization. Moreover, the transform sample derived from the original sample can weaken the effect, on the image, of illuminations. As we know, for visible light images, strong illuminations usually produce images with high gray values but weak illuminations will generate low gray values. As a result, for the same object, the appearance will vary with illuminations. The normalization process to produce the transform sample can effectively resist this influence and can partially weaken the original sample difference of the same object caused by various illuminations, which is beneficial to correct classification of images. On the other hand, because the original sample and transform sample are both reasonable representations of the image, to simultaneously exploit them for pattern classification is a feasible and proper way.

The second characteristic of the proposed method is that it simultaneously uses global and local dissimilarity measures of samples for classification. Moreover, the exploited three dissimilarity measures have enough difference. In particular, the center distance aims to evaluate global dissimilarity. It can be viewed as an average distance of the probe sample to a class. On contrary, the block distance and distance ratio are evaluations of local dissimilarity. The distance ratio is a special dissimilarity designed by us. It is calculated in the original sample space. It is not a conventional distance but a ratio, of the nearest distance of the probe sample to a class, to the sum of the nearest distances of the same probe sample to all other classes. The smaller this ratio is, the larger the probability of the probe sample being from the corresponding class is. In this sense, the distance ratio has the same meaning as a conventional distance. Thus, in our method it is reasonable to integrate the distance ratio with the other two distances for classification.

As a widely used local dissimilarity, the nearest distance provides an effective means to represent local space relationship between the probe sample and training samples and the corresponding nearest neighbor classifier can usually obtain relatively satisfactory result. However, it still has the following issue. As we know, for a real-world application, a limited number of available samples are observation results of a sampling procedure of the population in which all samples appear. Fig. 2 can partially illustrate this idea. In this figure, (a) is supposed to be the true sample space i.e. population. (b) and (c) stand for the results of two sampling procedures of the true sample space. Of course, both (b) and (c) contain only a part of the total samples that appear in (a). We know that different sampling procedures obtain different available samples, thereby they also generate different nearest distances for the same probe sample and same classes. Actually, since the nearest distance varies with the sampling procedure, it seems to be partially random from a probabilistic point of view. As a consequence, exploiting only the nearest distance to perform classification is not completely reliable. Thus, we propose to incorporate the global dissimilarity i.e. center distance with the nearest distance as well as distance ratio for classification of images.

The rationale of the second characteristic of the proposed method is also partly supported by numerical computation. Fig. 3 shows the used three categories of dissimilarity measures of samples from the Georgia Tech face database. In the case corresponding to this figure, the first 10 images of each person are used as gallery samples and the remaining images are exploited as probe samples. The figure depicts three categories of distance measures of the last probe with respect to all classes. We see that they all vary with the No. of the class and the variation trends are partially similar, but there still exist difference. Specifically, the Person correlation between the normalized distance ratio and normalized center distance is 0.74. The Person correlation between the normalized center distance and normalized block distance is 0.70. The Person correlation between the normalized distance ratio and normalized block distance is 0.93. The analyses above of course indicate that the three categories of dissimilarity measures used in our method contain enough difference and useful information for predicting label of the probe sample. As a result, combining them to perform classification is reasonable.

Furthermore, the simultaneous use of the three dissimilarities is reasonable also owing to the fact that both the original distance ratio and center distance are obtained on the basis of the Euclidean distance, whereas the nearest block distance is attained using a unique metric. We compare the block distance and Euclidean distance as follows. As we know, the widely used Euclidean distance calculates the nearest distance i.e. straight line distance between two points. However, the block distance between two points can be viewed as the reachable distance between two points located in a road network. In this sense, the block distance has clear physical meaning for measuring real-world distances. Fig. 4 visually shows difference of the block distance and Euclidean distance between two points in a two-dimensional space. First of all, for two points

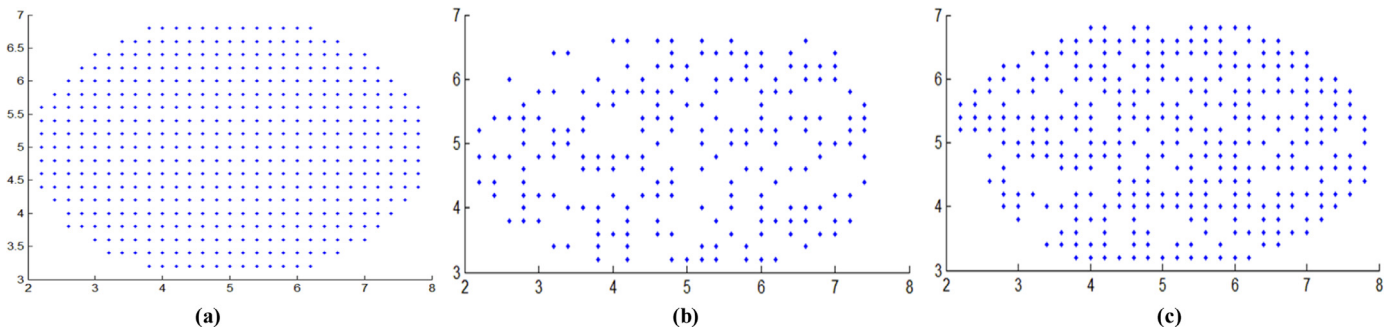

Fig. 2. Illustrations of the true sample space (population) i.e. (a) and the results of two sampling processes of the true sample space, i.e. (b) and (c). The sampling processes mean that both (b) and (c) contain only a part of the total samples that appear in (a).

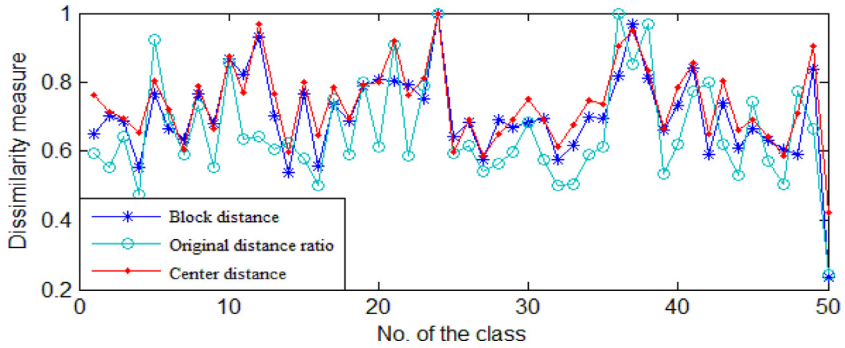

Fig. 3. Three categories of dissimilarity measures of samples on the GT face database. These three categories of dissimilarity measures are simultaneously used in our method.

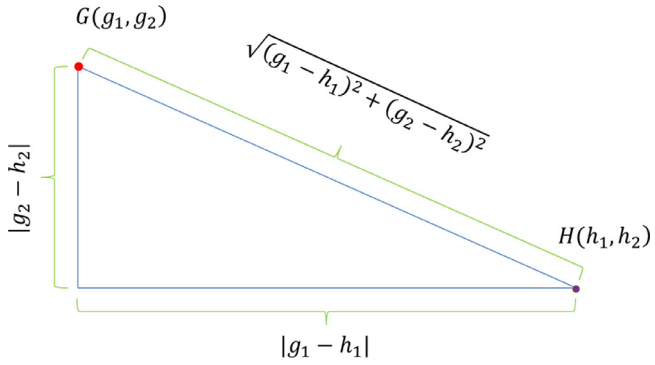

Fig. 4. Comparison of the block distance and Euclidean distance in a two-dimensional space. For points G and H, the block distance is $|g_1-h_1|+|g_2-h_2|$ and the Euclidean distance is $\sqrt{(g_1-h_1)^2+(g_2-h_2)^2}$. Moreover, $|g_1-h_1|$and $|g_2-h_2|$ are the lengths of the two sides of a right-angled triangle.$\sqrt{(g_1-h_1)^2+(g_2-h_2)^2}$ is the length of the hypotenuse of the same right-angled triangle. Based on rules of the triangle, we know that $|g_1-h_1|+|g_2-h_2| > \sqrt{(g_1-h_1)^2+(g_2-h_2)^2}$, so the block distance between two points must be not less than the Euclidean distance. Only under very special condition (i.e. either $|g_1-h_1| = 0$ or $|g_2-h_2| = 0$), they may be identical.

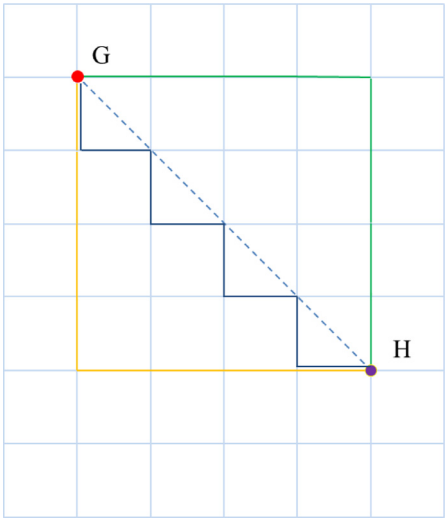

Fig. 5. Illustration of the block distance and Euclidean distance between two points in a two-dimensional space. It visually tells us that for real blocks and two points, maybe there are more than one reachable pathway from a point to another point. Specifically, the orange lines, green lines and dark blue lines depict three reachable pathways from point G to point H. However, the respective length sums of the three pathways are the same. The dotted line represents the Euclidean distance of the two points. In this case, though the Euclidean distance of the two points is the shortest distance, there is no corresponding real pathway.

in a two-dimensional space, the block distance is always equal to the length sum of the two sides of a right-angled triangle, whereas the Euclidean distance is identical to the length of the hypotenuse of the same right-angled triangle. Moreover, because the length sum of two sides of a triangle must be greater than the length of the third side, the block distance and Euclidean distance have the theoretical property that in a two-dimensional space the block distance between two points is almost always greater than the Euclidean distance.

Fig. 5 visually indicates that the block distance between two points is the reachable distance between two points located in a road network. All lines in this figure denote available roads whereas all other locations are not reachable. This figure visually tells us that for real blocks and two points, maybe there are more than one reachable pathway from a point to another point. However, the length sum of each reachable pathway is the same. Fig. 5 also shows that the Euclidean distance between two points is indeed the shortest distance and there is only a sole "pathway" corresponding to this distance. However, for a road network, this sole "pathway" is an imaginary road and is not real. In other words, it is unreachable.

The fact that the block distance between two samples is usually greater than the Euclidean distance also has the following indicative meaning. When the distance between sample vectors s and t is calculated, we call the s-t difference vector. It is known that main difference of conventional distance metrics is that they use various ways to cope with the difference vector. Compared with the Euclidean distance, the block distance allows all components of the difference vector to be equally dealt with. In other words, since the absolute values of all components of the difference vector are directly summed, the obtained distance faithfully reflects the magnitudes of all components. However, the Euclidean distance is directly associated with the squares of all components of the difference vector, so it enhances the effect of the components with large absolute values and weakens the effect of the components with small absolute values. For example, if the first and second components of a two-dimensional difference vector is −0.1 and 0.85, then the block distance is 0.95. However, in calculation of the Euclidean distance, the squares of −0.1 is just 0.01, much less than the squares of 0.85, so the effect of the first component is almost neglectable and the final Euclidean distance is 0.856, which is almost completely dominated by the second component of the difference vector. In real-world applications, it is possible that the components with large absolute values contain more noise. As a consequence, the Euclidean distance will cause much more noise than the block distance. For example, for an occluded face image and a normal face image of the same person, the components of

Table 1
Rates of classification errors (%) of different methods on the GT face database.

	6	7	8	9	10	11
Our method	**30.2**	**27.5**	**24.6**	**21.3**	**20.0**	**19.0**
NNC in the original space	35.8	32.3	29.7	26.0	27.6	24.0
NNC in the transform space	40.7	37.3	34.6	32.0	33.2	29.5
NNCC in the transform space	41.8	38.3	37.1	36.0	32.4	30.5
NNBC in the transform space	32.9	29.5	27.4	26.0	24.0	22.5
Collaborative representation	44.4	41.5	40.6	39.0	36.0	32.5
L1LS	50.4	49.5	47.4	49.7	48.4	48.0
FISTA	43.8	45.0	44.9	46.7	43.2	46.5

the difference vector corresponding to the occluded facial region must have large absolute values. As a consequence, the Euclidean distance will enhance the noise caused by the occlusion but the block distance will not. On the other hand, for the difference vector of two face images respectively from two persons, some components of the difference vector have large absolute values and simultaneously might reflect essential differences of the two faces, so the amplifying effect of the Euclidean distance is also useful for distinguishing the faces. Therefore, integrating the block distance and Euclidean distance is a proper way to measure the difference of face images under various circumstances.

The proposed method can be viewed as a fusion method. However, it differs from conventional fusion methods, which fuse information at the feature level, score level or decision levels. In conventional fusion methods, usually at least two categories of data are available. For example, multi-biometrics exploits different biometrics traits such as faces, palmprint and fingerprint at feature level, score level or decision level to generate the final recognition result. In our proposed method, only one category of data i.e. original image data are available and fusion is carried out on the basis of the three categories of dissimilarity metrics on the same data.

As we know, for a multi-view problem with three views and a single dissimilarity metric, a multi-view learning method also can get three dissimilarity values to represent relationship of each pair of samples. In this sense, our method is partially equivalent to a multi-view learning method. In other words, we may say that the three dissimilarity metrics in our method allows the single modal data to be observed from three points of view.

4. Experimental results and analyses

We present experimental results and necessary analyses in this section. We call the nearest neighbor classification NNC. The nearest neighbor classification implemented in the original space and transform space are called NNC in the original space and NNC in the transform space, respectively. The nearest neighbor center-distance based classification is referred to as NNCC. The nearest neighbor block-distance based classification is referred to as NNBC. Besides experiments on the above methods, we also compare our method with the widely used collaborative representation method and two well-known sparse representation algorithms, L1LS [11] and the fast iterative shrinkage thresholding algorithm (FISTA) [1], as they are three classification methods on basis of a same special dissimilarity measure and classification rule. All methods except for collaborative representation, L1LS and FISTA exploit the nearest neighbor classification strategy, therefore the experimental comparison is fair. When testing our method, we adopt the same parameter setting $r_1 = 0.1$, $r_2 = 0.1$ and $r_3 = 0.8$ for the original Georgia Tech and ORL face databases. In the experiment on the AR database which contains occluded face images, $r_1 = 0.45$, $r_2 = 0.1$ and $r_3 = 0.45$. We choose the parameter values based on our empirical experience. In real-world applications, one may select proper values for the parameters via a validation dataset. In particular, if the parameters are set to different values, the validation dataset will attain different accuracies. The parameter values corresponding to the highest accuracy can be selected as the optimal values. For all experimental results shown in the following tables, smaller numbers means better performance, because what they mean are rates of classification errors rather than accuracies. For all tables, the first row shows how many facial images of a person are used as gallery samples. For example, number 6 in the first row of a table means that the first six images of every person are exploited as gallery samples and its other images are used as probe samples. The code of the designed method will be accessible at http://www.yongxu.org/lunwen.html.

4.1. On the Georgia Tech (GT) face database

The GT face database includes 750 facial images captured from fifty persons. As previous studies usually did, we resize each facial image into a size of 40 by 30 grey pixels. Variations in poses, illuminations and facial expressions are reflected by different images of the same person [8]. Table 1 provides rates of classification errors of different methods on the GT face database. It indicates that

Table 2
Rates of classification errors (%) of different methods on the ORL face database.

	2	3	4	5	6	7
Our method	**15.0**	**12.5**	**10.0**	**7.0**	**3.1**	**4.2**
NNC in the original space	17.5	13.6	10.8	9.0	5.0	6.7
NNC in the transform space	20.3	17.5	12.9	11.5	6.9	10.8
NNCC in the transform space	21.9	21.1	20.0	17.0	15.6	**4.2**
NNBC in the transform space	16.3	14.6	11.7	7.5	3.8	**4.2**
Collaborative representation	16.6	13.9	10.8	11.5	8.1	8.3
L1LS	20.0	18.9	14.6	13.5	11.9	6.7
FISTA	18.4	16.8	12.1	13.5	14.4	10.8

our method not only is better that NNC in the original space, NNC in the transform space, NNCC in the transform space and NNBC in the transform space, but also obtains fewer errors than collaborative representation, L1LS and FISTA. For instance, when the first 6 images of each face are used as gallery samples and the others are exploited as probe samples, the rate of classification errors of our method is only 30.2%, much smaller than those of NNC in the original space and the other methods.

4.2. On the ORL face database

The face images of the ORL face database are collected from forty persons and 10 face images are available for each person [17]. In this database, besides variations in poses, illuminations and facial expressions are conveyed, the images also have difference in glasses. In other words, some faces are imaged with glasses but the others are not. Table 2 shows that the rate of classification errors of our method is lower than those of all other methods.

4.3. On the AR face database

The used AR face database contains face images of 120 persons. In this database, for every person, there are 26 available face images. A remarkable characteristic of this database is that a number of face images convey illumination changes or severe occlusion

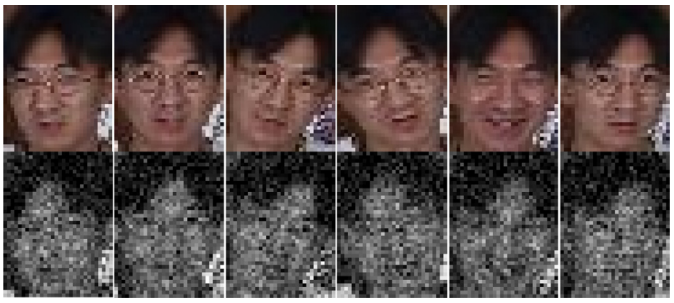

Fig. 6. Six original face images (shown in the first row) of the same person from the GT face database and the corresponding corrupted face images (shown in the second row).

[15]. Table 3 offers rates of classification errors of different methods on the AR face database. We also see that our method always outperforms all other methods in terms of the rate of classification errors. In some cases, the rate of classification errors of our method is lower than one half of those of some other methods. In this sense, our method is very effective for recognition of faces under the condition that some faces are partially occluded.

4.4. On the GT corrupted face database

In order to see how different methods are sensitive to noise, we conduct this experiment. We impose Gaussian noise to each original face image in the GT database by using Matlab function "imnoise" and set type of noise as "gaussian". Moreover, when using this function, we set the mean and variance to zero and 0.01 respectively. Fig. 6 provides six original face images of the same person and the corresponding corrupted face images. We present the experimental result in Table 4. The parameters of our method are set to $r_1 = 0.6$, $r_2 = 0.3$ and $r_3 = 0.1$. We see again that our method attains lower rate of classification errors than all other methods and our method achieves very satisfactory performance. In addition, we find that in the transform space in this experiment, both the block distance and nearest distance perform worse than the

Table 3
Rates of classification errors (%) of different methods on the AR face database.

	3	4	5	6	7	8
Our method	**23.0**	**22.0**	**17.7**	**15.0**	**13.3**	**12.1**
NNC in the original space	35.0	34.2	27.9	23.8	22.6	21.3
NNC in the transform space	40.9	42.7	40.3	38.9	40.2	41.8
NNCC in the transform space	41.4	42.7	42.1	42.1	43.3	44.2
NNBC in the transform space	38.2	40.0	37.3	35.2	35.3	36.8
Collaborative representation	31.4	32.5	30.4	29.2	29.7	30.1
L1LS	34.1	35.8	35.2	34.8	38.3	38.2
FISTA	38.7	41.3	41.7	42.3	45.4	46.5

Table 4
Rates of classification errors (%) of different methods on the GT corrupted face database.

	6	7	8	9	10	11
Our method	**32.7**	**29.5**	**26.9**	**25.7**	**24.4**	**21.0**
NNC in the original space	37.6	35.3	32.0	29.7	28.8	25.5
NNC in the transform space	59.1	57.8	56.3	53.7	53.6	49.0
NNCC in the transform space	45.3	40.3	39.4	39.3	36.8	33.5
NNBC in the transform space	56.0	52.0	49.7	47.3	48.0	45.5
Collaborative representation	58.0	55.0	52.3	52.7	58.4	52.0
L1LS	60.4	59.5	54.6	56.0	61.6	57.5
FISTA	53.6	51.0	56.0	57.3	52.0	56.0

center distance when they are exploited to classify probe face images. This partially means that under a severe noisy circumstance the local dissimilarity measure is not so reliable in representing the class relationship of images as the global dissimilarity measure, which is consistent with the analysis offered in Section 3.

5. Conclusion

In this paper, we provide a simple method to combine global and local dissimilarity measures from two sample spaces to perform image classification. This method not only is easy to implement but also has obvious rationale. Its good performance implies that for classification of high-dimensional image samples, feature extraction and classifier design may not be necessary. The proposed simultaneous use of the original distance ratio from the original sample space as well as the global class distance and nearest block distance from the transform sample space, is very suitable for the classification problem of the high-dimensional image samples. For a high-dimensional space, a limited number of available samples are indeed sampling results of the population. As a consequence, the sole use of the nearest neighbor distance might not well represent the dissimilarity relation of the probe sample and each class and combining the global class distance and local distances is a good way. A remarkable advantage of the nearest block distance is that it properly and equally treats all components of the difference vector. The proposed novel distance ratio is not a conventional distance metric but has similar meaning. Moreover, it provides complementary information for dissimilarity measures of samples. The experimental results demonstrate that the proposed method does achieve a very high accuracy in comparison with the other methods. They also show that in conventional cases the block distance can usually perform very well in identifying the class relationship of probe images whereas under severe noisy circumstance the center distance can attain relatively low rate of classification errors in comparison with the block distance. In the future, we will try to improve the proposed method. For example, we will study whether the proposed dissimilarity may be replaced by others such as the kernel distance.

Declaration of Competing Interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled "Combining dissimilarity measures for image classification".

Acknowledgments

This paper is partially supported by Economic, Trade and Information Commission of Shenzhen Municipality (Grant No. 20170504160426188) and Shenzhen Municipal Science and Technology Innovation Council (Grant Nos. GJHZ20180419190732022 and JCYJ20180306172101694).

References

[1] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (1) (2009) 183–202.
[2] R. Benouini, I. Batioua, K. Zenkouar, A. Zahi, H. El Fadili, H. Qjidaa, Fast and accurate computation of Racah moment invariants for image classification, Pattern Recognit. 91 (2019) 100–110.
[3] N. Biswas, S. Chakraborty, S.S. Mullick, S. Das, A parameter independent fuzzy weighted k-nearest neighbor classifier, Pattern Recognit. Lett. 101 (2018) 80–87.
[4] H. Cao, S. Bernard, R. Sabourin, L. Heutte, Random forest dissimilarity based multi-view learning for Radiomics application, Pattern Recognit. 88 (2019) 185–197.
[5] Z. Chen, X.J. Wu, J. Kittler, A sparse regularized nuclear norm based matrix regression for face recognition with contiguous occlusion, Pattern Recognit. Lett. (2019), doi:10.1016/j.patrec.2019.05.018.
[6] H. Chi, H. Xia, L. Zhang, C. Zhang, X. Tang, Competitive and collaborative representation for classification, Pattern Recognit. Lett. (2018), doi:10.1016/j.patrec.2018.06.019.
[7] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Second ed., John Wiley & Sons, New York, 2000.
[8] N. Goel, G. Bebis, A. Nefian, Face recognition experiments with random projection, in: Processing of the Biometric Technology for Human Identification II. International Society for Optics and Photonics, 5779, SPIE, Bellingham, WA, USA, 2005, pp. 426–437.

[9] B.K. Iwana, V. Frinken, K. Riesen, S. Uchida, Efficient temporal pattern recognition by means of dissimilarity space embedding with discriminative prototypes, Pattern Recognit. 64 (2017) 268–276.
[10] X.Y. Jing, S. Li, D. Zhang, J. Yang, J.Y. Yang, Supervised and unsupervised parallel subspace learning for large-scale image recognition, IEEE Trans. Circuits Systs. Video Technol. 22 (10) (2012) 1497–1511.
[11] S.J. Kim, K. Koh, M. Lustig, S. Boyd, D. Gorinevsky, An interior-point method for large-scale l1-Regularized least squares, IEEE J. Sel. Top. Signal Process 1 (4) (2007) 606–617.
[12] S. Li, Y. Fu, Learning robust and discriminative subspace with low-rank constraints, IEEE Trans. Neural. Netw. Learn. Syst. 27 (11) (2015) 2160–2173.
[13] J. Li, B. Zhang, G. Lu, D. Zhang, Generative multi-view and multi-feature learning for classification, Information Fusion 45 (2019) 215–226.
[14] L. Liu, S. Li, C.P. Chen, Iterative relaxed collaborative representation with adaptive weights learning for noise robust face hallucination, IEEE Trans. Circuits Syst. Video Technol. 29 (5) (2018) 1284–1295.
[15] Martinez A.M., Benavente R., 1998. The AR face database. CVC technical report.
[16] Y. Peng, L. Li, S. Liu, X. Wang, J. Li, Weighted constraint based dictionary learning for image classification, Pattern Recogn. Lett. (2018).
[17] F.S. Samaria, A.C. Harter, Parameterisation of a stochastic model for human face identification, in: Processing of the 1994 IEEE Workshop on Applications of Computer Vision, IEEE, Piscataway, NJ, USA, 1994, pp. 138–142.
[18] T. Shu, B. Zhang, Y.Y. Tang, Multi-View classification via a fast and effective multi-view nearest-subspace classifier, IEEE Access 7 (2019) 49669–49679.
[19] A. Skowron, H. Wang, A. Wojna, J. Bazan, Multimodal classification: Case studies, in: F.P. James, S. Andrzej (Eds.), Transactions on Rough Sets V, Springer, Berlin, Heidelberg, 2006, pp. 224–239.
[20] J. Song, X. Xie, G. Shi, W. Dong, Multi-layer discriminative dictionary learning with locality constraint for image classification, Pattern Recognit. 91 (2019) 135–146.
[21] Y. Tao, J. Yang, W. Gui, Robust $l_{2,1}$ norm-based sparse dictionary coding regularization of homogenous and heterogenous graph embeddings for image classifications, Neural Process. Lett. 47 (3) (2018) 1149–1175.
[22] Q. Tuo, H. Zhao, Q. Hu, Hierarchical feature selection with subtree based graph regularization, Knowl. Based Syst. 163 (2019) 996–1008.
[23] C. Wang, Q. He, M. Shao, Q. Hu, Feature selection based on maximal neighborhood discernibility, Int. J. Mach. Learn. Cyb. 9 (11) (2018) 1929–1940.
[24] J. Wen, Y. Xu, Z. Li, Z. Ma, Y. Xu, Inter-class sparsity based discriminative least square regression, Neural Networks 102 (2018) 36–47.
[25] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2008) 210–227.
[26] Y. Xu, D. Zhang, J. Yang, J.Y. Yang, A two-phase test sample sparse representation method for use with face recognition, IEEE Trans. Circuits Syst. Video Technol. 21 (9) (2011) 1255–1262.
[27] Y. Xu, Q. Zhu, Z. Fan, M. Qiu, Y. Chen, H. Liu, Coarse to fine k nearest neighbor classifier, Pattern Recognit. Lett. 34 (9) (2013) 980–986.
[28] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: which helps face recognition? in: Processing of the 2011 International conference on computer vision, Piscataway, NJ, USA, IEEE, 2011, pp. 471–478.
[29] X. Zhang, J. Sun, S. Ma, Z. Lin, J. Zhang, S. Wang, W. Gao, Globally variance–constrained sparse representation and its application in image set coding, IEEE Trans. Image Process. 27 (8) (2018) 3753–3765.
[30] C. Zheng, N. Wang, Collaborative representation with k-nearest classes for classification, Pattern Recognit. Lett. 117 (2019) 30–36.
[31] G. Zhong, Y. Zheng, S. Li, Y. Fu, Slmoml: online metric learning with global convergence, IEEE Trans. Circuits Syst. Video Technol. 28 (10) (2017) 2460–2472.
[32] W. Zhong, L. Jiang, T. Zhang, J. Ji, H. Xiong, Combining multilevel feature extraction and multi-loss learning for person re-identification, Neurocomputing 334 (2019) 68–78.
[33] W. Zuo, F. Wang, D. Zhang, L. Lin, Y. Huang, D. Meng, L. Zhang, Distance metric learning via iterated support vector machines, IEEE Trans. Image Process. 26 (10) (2017) 4937–4950.