

Received September 7, 2020, accepted September 23, 2020, date of publication September 28, 2020,  
date of current version October 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3027311

# Fuzzy Granule Manifold Alignment Preserving Local Topology

WEI LI<sup>1</sup>, JIANWU XUE<sup>1</sup>, YUMIN CHEN<sup>1</sup>, (Member, IEEE), XUEBAI ZHANG<sup>1</sup>, CHAO TANG<sup>2</sup>,  
QIANG ZHANG<sup>1</sup>, AND YIFANG GAO<sup>1</sup>

<sup>1</sup>School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China

<sup>2</sup>School of Artificial Intelligence and Big Data, Hefei University, Hefei 230601, China

Corresponding author: Wei Li (drweili@hotmail.com)

This work was supported in part by the Science and Technology Planning Guidance Project, Xiamen, China, under Grant 3502Z20179038, in part by the National Natural Science Foundation of China under Grant 61976183, in part by the Natural Science Foundation of Fujian Province, China, under Grant 2019J01850, in part by the Key Teaching and Research Project of Hefei University of China under Grant 2018 hfjyxm09, in part by the Natural Science Foundation of Anhui Province of China under Grant 2008085MF202, and in part by the Xiamen University of Technology, China, under Grant YKJ19004R.

**ABSTRACT** Granular computing has the advantage of discovering complex data knowledge, and manifold alignment has proven of great value in a lot of areas of machine learning. We propose a novel algorithm of fuzzy granule manifold alignment (FGMA), where we define some new operations, measurements, and local topology of fuzzy granular vectors in fuzzy granular space. Furthermore, the algorithm is very different from Semi-supervised and Procrustes algorithm because predetermining correspondence is not necessary. A projection is learned that can map instances described by two types of features to a low-dimensional space. Meanwhile, the local topology of the fuzzy granular vector induced by the instance is also preserved and matched within each set in lower dimensional space. This approach makes it possible to directly compare between data instances in different spaces. We convert an alignment problem of data in feature space into fuzzy granular manifold alignment problem of granular space. Specifically, we first define fuzzy granule, fuzzy granular vector, operations, and measurements in fuzzy granular space and gave proofs of theorems and deductions. Next, the local topology around the fuzzy granular vector is introduced and the optimal local topology matching can be achieved by minimizing their Frobenius norm. Finally, two manifolds are connected and the optimal mapping can be calculated to obtain dimensionality reduction of the joint structure. Thus, the corresponding relationship between two data instances can be got. We verified this algorithm in Oxford image and Alzheimer's disease voice dataset. Theoretical analysis and experiments demonstrate the algorithm proposed is robust and effective.

**INDEX TERMS** Granular computing, fuzzy sets, manifold learning, alignment, local topology.

## I. INTRODUCTION

The term manifold learning was first proposed by Bregler and Omohundro in 1995 [1], [2]. It refers to restoring the underlying low-dimensional manifold structure on the basis of high-dimensional data, while building a low-dimensional description of data. This low-dimensional description reflects intrinsic variables that control the changes in the distribution of manifold data, that is, the minimum independent variables require to describe and perceive the changes in the distribution of data points on the manifold. In a broad sense, as long

The associate editor coordinating the review of this manuscript and approving it for publication was Bilal Alatas.

as it is based on the manifold distribution assumption of the dataset, any feature learning method for the purpose of learning the internal rules and structural characteristics of the dataset can be regarded as the category of manifold learning. The difference between various manifold learning methods is limited to their different ways of collecting neighborhood information and the structural characteristics of the neighborhood. Manifold learning can be classified into global and local methods [3]. The global method is to maintain the distance relationship of all data instance points at various scales from the perspective of the global geometry of the distribution structure of the dataset. That is, when constructing a low-dimensional embedded representation, we need to

project points on manifold that are close to the distribution to neighboring points in lower dimensional space. Meanwhile, data points that are distributed farther are mapped into lower dimensional space, and their distribution relationship at a long distance is also maintained. Representative algorithms include ISOMAP and Maximum Variance Unfolding (MVU) and Diffusion Maps (DM). This idea based on global geometric structure is simple and easy to understand, and can give an accurate description of the global distribution structure of data. However, this type of method generally requires the subset of low-dimensional space to be convex, otherwise the geometric distance between the data points cannot be accurately described. In addition, because the distance between any pair of data needs to be calculated, the calculation of the algorithm is expensive. The local method is very different from the global one. It uses the perspective of local geometry to assure data points that are close within local neighborhood obtain similar projection positions on the low-dimensional space. This type of method constructs a low-dimensional description by maintaining the change of the distribution structure of the manifold data in the local neighborhood. Although it is impossible to accurately describe the overall structure of the data, its calculation cost is low. And for complex manifold structures, there is no correct priori knowledge as the guidance, so the calculation of the distance between distant points is often not guaranteed to be precision. Therefore, local methods are more widely employed in practice. Representative algorithms involve local tangent space alignment (LTSA), Laplacian eigenmaps (LE), local linear embedding (LLE) etc.

## II. RELATED WORK

Three papers published in Science in 2000 brought manifold learning to a new stage of research [4]–[6]. In these three papers, not only was the biological basis of the existence of perceptual manifolds demonstrated, but also the isometric feature mapping algorithm and the local linear embedding algorithm were proposed, and good results were obtained. The method ISOMAP maintains the geodesic distance between data points and is successfully used in biomedical data visualization [7], [8]. In the LLE method, data points can be denoted by a linear combination of their neighbors, so local properties of data are maintained in dimensionality reduction, and they are successfully applied to non-convex manifolds, but the artificial biomedical database cannot be visualized [9]. Another Laplacian Eigenmaps (LE) algorithm that has attracted attention considers the distance between neighbors, and obtains a low-dimensional data representation by maintaining the local properties on the manifold [9], [10]. Although the algorithms mentioned above have excellent performance in dimensionality reduction, they do not get an explicit mapping. When new data is input, it cannot extract features fast and effectively. To obtain solution of this problem, Locality Preserving Projections (LPP) algorithm [11]–[13], Neighborhood Preserving Embedding (NPE) algorithm [14], [15] and Orthogonal Neighborhood Preserving Projections (ONPP) algorithm [16]

were successively proposed. These algorithms have similar ideas. First, they establish a neighbor graph for the instance points of the original dataset and construct a suitable relationship matrix to characterize the similarity of the pair of points inside the neighbor graph. Then these algorithms retain the neighbor relationship of the original dataset when reducing the dimension. And finally the linear embedding map of the explicit form through the optimization criterion can be obtained. This type of method can map new points directly into a lower dimensional space by linear embedding mapping. Whereas the methods do not adopt the category information in the dataset. The separability between the projected categories is not prominent. Hence the methods are unfit for classification and recognition of data, but suitable for dimensionality reduction or clustering [17]. Peng and his colleagues proposed a novel subspace clustering approach by introducing a new deep model-Structured AutoEncoder (StructAE) to handle realistic data without the linear subspace structure in 2018 [18]. Next, they proposed a novel objective function to project raw data into one space in which the projection embraces the geometric consistency (GC) and the cluster assignment consistency (CAC) [19]. Furthermore, they proposed a novel clustering method by minimizing the discrepancy between pairwise instance assignments for each data point [20]. These researches were very important to align instances.

Granular computing has the superiority of discovering complex data knowledge. If manifold alignment is combined with granular computing, a new algorithm can be designed to enhance the performance. The idea of Granular computing is derived from concept of fuzzy information granulation proposed by Zadeh [21]. He thought that there are three main characteristics of human cognition, namely, granulation (breaking wholes into parts), organization (combining parts into wholes), and causation [22]. Hobbs presented the concept of Granularity in 1985 [23]. Then, Lin presented the concept of granular computing on the basis of binary relations in 1998 [24]. Yager *et al.* also talked over importance of granular computing in intelligent engineering in the same year [25]. In 2000, Yao gave the construction and calculation of granules in granular computing in detail [26]. In 2002, he studied the problem of information granulation and concept approximate structure in the rough set theory [27]. After that, Yao concentrated on basic concept of granular computing and gave a granular computing model via region division using set theory [28]; analyzed the past, present and future development of granular computing and presented triadic theory [29]; deeply discussed connection between cognitive science and granular computing and constructed a granular computing framework for cognitive concept learning [30]. The rough outline of the granular computing gradually forms. Pedrycz *et al.* published the book “Handbook of granular computing”, which aims to provide guidance and assistance in granular computing related to computing intelligence, and so on [31]. Ling Zhang and Bo Zhang published the book on problem solution of quotient space, which elaborated a

mathematical model of hierarchical multi-granule computing based on quotient space [32]. Duoqian Miao, Deyi Li, Yiyu Yao *et al.* introduced the process of research on the uncertainty of granular computing and the cross-over study in [33].

Wang and his colleagues analyzed and summarized the uncertainty problems on granular computing models [34]. Liang *et al.* combined concept of information granule with entropy theory, and established the complementary relationship between them [35]. Yao *et al.* gave a new way of thinking about prospects and challenges of future granular computing in [36]. Wu *et al.* established a rough approximation model of the order granular structure in the order information system of multi-granularity marking [37]. For the three core issues of information granulation, information granularity and granular operation under granular computing, Qian *et al.* established a unified representation framework for granular computing based on sets [38]. Hu *et al.* realized the attribute reduction of mixed data by constructing a new information granulation method on fuzzy rough sets [39]. Xu *et al.* established a formal conceptual system describing human cognitive processes, and gave definitions and algorithms for granular transformation of related information [40]. Herbert and Yao regard the growth and absorption of neurons as the construction and decomposition of granules, and then proposed a hierarchical self-organizing mapping granular computing framework for neural networks [41]. Jankowski *et al.* used the interactive operation between complex information granules in rough set model to present an approach to deal with uncertain problems in complex systems [42]. Dubois and Henri discussed the common features between extended fuzzy sets from two aspects of similarity relationship and formal concept [43]. Peters and Weber summarized and proposed a unified framework for existing dynamic granule clustering algorithms [44]. Li *et al.* proposed formal concept learning method on the basis of granular computing from angle of cognitive computing [45]. Xu *et al.* presented a granular computing method that uses information granules to describe machine learning for fuzzy datasets [46]. Chiaselotti *et al.* connected the automorphism of graphs with the indistinguishable relationship, and introduced the relationship between simple graphs based on adjacency matrices and granular computing [47]. Salehi *et al.* conducted a systematic classification study on granular computing related research from four perspectives of key areas, contribution types, research types and research frameworks. They found that the research on clustering analysis for information granules was rare, and compared the effect of information granulation of five common clustering algorithms [48]. Al-Hmouz *et al.* introduced a granular computing framework on the basis of time series description and prediction [49]. Wang and his colleagues surveyed existing research of granular computing from optimization and conversion of granularity and multi-granularity joint problem solving, and presented graph for relationship among three primary modes of granular computing [50]. Xu and his colleagues presented a

local-density-based optimal granulation model [51]. Granular computing has achieved lots of results in theory, methods and applications, but there are still many key problems to be solved in the analysis and processing of big data, such as the abstract description of multi-level granular space and knowledge acquisition technology, the establishment and optimization of multi-granularity model, the interpretation and construction of the multi-level granularity spatial structure, the conversion mechanism and method between the granular layer and the granularity, etc. And three-way decision proposed by Yiyu Yao is one of good methods to find these solutions. In addition, Li *et al.* presented a fastest robust path optimization algorithm to form the best routing by combining the effect of traffic events [52]. Li and his colleagues adopted granular computing to design boosted  $k$  nearest neighbor classifiers [53] and search voice encrypted scheme [54]. Mencar and Pedrycz proposed the definition of granular counting, which realized in the presence of uncertain data modeled through possibility distributions [55] and this is also a new angle to study uncertain data.

### III. CONTRIBUTIONS

Our contributions are as follows:

- We convert an alignment problem of instance in feature space into fuzzy granule manifold alignment problem in fuzzy granular space to solve. This method incorporates the idea of granular computing and solves problems from different levels. This method can observe and process data sets hierarchically from different granularity, which reduces the complexity of alignment problem.
- We propose the term fuzzy granular vector by defining new operations and metrics and design the algorithm of fuzzy granule manifold alignment based on them. Moreover, this algorithm can align instances without predetermining correspondence, which is required for Semi-supervised and Procrustes algorithm.

### IV. THE PROBLEM

As defined in Table 1, let  $P$  be a set of instances sampled from manifold  $\mathcal{P}$ , and let  $Q$  be a set of instances collected from manifold  $\mathcal{Q}$ . Learning mappings  $\alpha$  and  $\beta$  to a new space  $Z$  is our aim. Also, we need to preserve the neighborhood relationships inside of  $P$  and those of  $Q$  in new space  $Z$  respectively. Suppose that local topology of  $p_i \in P$  and  $q_j \in Q$  can be well matched in the original spaces, then  $p_i$  and  $q_j$  will also be matched in the new space  $Z$ . Our goal is to align the instances in the datasets  $P$  and  $Q$ . Since the instances in these two datasets are represented by features of different dimensionality, it is very difficult to directly compare  $p_i \in P$  and  $q_j \in Q$ . In order to establish the relationship between them, we use fuzzy granular vector constructed by  $p_i$  and fuzzy granular vectors induced by its neighbors to jointly characterize the local topology of  $p_i$ . Similarly, the local topology of  $q_j$  is also expressed via this method.  $p_i$  and  $q_j$  may be directly compared by local topology denoted relations

TABLE 1. Notation.

Symbol	Notation
$\mathcal{P}, \mathcal{Q}$	Manifold
$P, Q, Z$	instance set
$p_i, q_j$	Point of instance set
$R, A, E$	Attribute set
$r_i, a_j, e_k$	Atom attribute or element in attribute set
$f_*(*)$	Value of point on atom attribute
$N_*(*)$	Fuzzy granule of point on atom attribute
$ N_*(*) $	Cardinality of fuzzy granule on atom attribute
$\hat{N}_*(*)$	fuzzy granular vector of point on atom attribute
$ \hat{N}_*(*) $	Module between fuzzy granular vectors on attribute set
$s_x(*, *)$	Distance between points on atom attribute
$\hat{s}_x(*, *)$	Distance between fuzzy granular vectors on attribute set

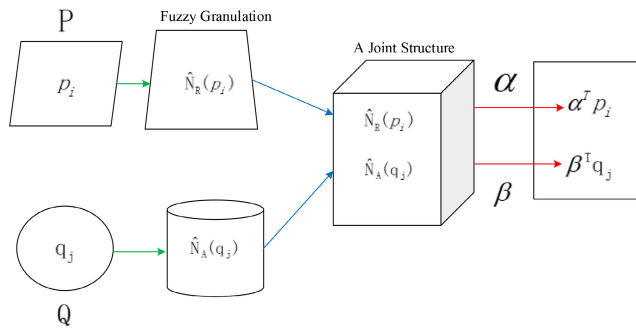


FIGURE 1. Illustration of the main algorithm.

instead of features. But  $p_i$  might be similar to many instances in  $Q$ . Therefore, to find which instance in  $Q$  is true similar is very difficult. As matter of fact, there are often lots of true match for adhibitions. Interestingly, discovering true match may be more difficult than solving the coupled problem of the original problem. The reason lies in preserving structures of two manifolds during matching process. So it can help us filter out some false positive matches. In the method, we firstly establish all likely matches for every instance based on local topology mentioned above. After that instance alignment problem is transformed into an embedding problem with constraint conditions, which can be found solution using generalized eigenvalue decomposition.

The algorithm flow is shown in Figure 1.  $p_i \in P$  and  $q_j \in Q$  come from varied manifolds, so they are not to be indirectly compared. The algorithm proposed learns the mappings  $\alpha$  and  $\beta$  which can map the instances from  $P$  and  $Q$ , respectively, to a new same space. In this way, samples from varied manifolds with similar local topology can be projected to similar locations and manifold structure can also be kept. In this process, first, each instance in  $P$  and  $Q$  is fuzzy granulated to construct a fuzzy granular vector. After that we need to analyze local topology of each fuzzy granular vector, put the fuzzy granular vectors and their local topology from  $P$  and  $Q$  into the joint structure, and find the best local topology match, so as to get the mapping  $\alpha$  and  $\beta$ . Thus,  $\alpha^T p_i$  and  $\beta^T q_j$  are in the same space and have the same dimensionality, which can be directly compared and be as the evaluation of the similarity between  $p_i$  and  $q_j$ .

## V. THE MAIN ALGORITHM

Given two instance sets  $\{P, R\}, \{Q, E\}$ , where  $P = \{p_1, p_2, \dots, p_m\}$ , its attribute set  $R = \{r_1, r_2, \dots, r_t\}$  and  $Q = \{q_1, q_2, \dots, q_n\}$ , its attribute set  $E = \{e_1, e_2, \dots, e_s\}$ . Let  $p_i \in P$  be a point defined in  $t$ -dimensional manifold  $\mathcal{P}$ . Its feature vector is  $(f_{r_1}(p_i), \dots, f_{r_t}(p_i))^T$ , and  $P$  is a  $t \times m$  matrix. Similarly,  $q_j$  is a point defined in a  $s$ -dimensional manifold  $\mathcal{Q}$ , its  $s$ -dimensional feature is  $(f_{e_1}(q_j), \dots, f_{e_s}(q_j))^T$ , and  $Q$  is a  $s \times n$  matrix. The aim is to discover the match point pair between  $P$  and  $Q$ . We first fuzzily granulate each point in  $P$  and  $Q$ . Then we can match the points according to the local topology of fuzzy granular vectors induced by them in the fuzzy granular space. Further these points from  $P$  and  $Q$  can be aligned or these matching point pairs can be found.

### A. CONVERTING DATA INTO FUZZY GRANULES

In general, it is fuzzy that the granularity of human reasoning and concept building. Fuzzy information granulation is generally achieved by fuzzy binary relationship defined in fuzzy granular space. Atomic attribute is fuzzy granulated and then we can construct fuzzy granular vectors. For  $\forall p_i, p_j \in P$  and  $\forall r \in R$ , the distance on attribute  $r$  between  $p_i$  and  $p_j$  can be denoted as:

$$s_r(p_i, p_j) = |f_r(p_i) - f_r(p_j)| \quad (1)$$

where  $s_r(p_i, p_j) \in [0, 1]$ .  $f_r(p_i)$  and  $f_r(p_j)$  express normalized value of  $p_i$  and  $p_j$  on the attribute  $r$ , respectively. According to distance definition, fuzzy granule can be constructed. For  $\forall p_i \in P, \forall r \in R$ , fuzzy granule of instance  $p_i$  on attribute  $r$  can be denoted as:

$$\begin{aligned} N_r(p_i) &= \frac{s_r(p_i, p_1)}{p_1} + \frac{s_r(p_i, p_2)}{p_2} + \dots + \frac{s_r(p_i, p_m)}{p_m} \\ &= \sum_{j=1}^m \frac{s_r(p_i, p_j)}{p_j} \end{aligned} \quad (2)$$

Here  $N_r(p_i)$  is the fuzzy granule induced by  $p_i$  (also seen as fuzzy neighborhoods of  $p_i$ ). “+” represents the union operator, and “-” denotes separator. This fuzzy granule can also be seen as a set composed of the distance pairs between instances. By accumulating the elements inside fuzzy granule, the cardinal of fuzzy granule is denoted as:

$$|N_r(p)| = \sum_{v \in P} s_r(p, v) \quad (3)$$

We have  $0 \leq |N_r(p)| \leq |P|$ . Here  $|P|$  represents the number of instance set  $P$ . Since an attribute can induce one fuzzy granule of one instance, one attribute set can induce one fuzzy granular vector of one instance. In other words, a fuzzy granular vector consists of many fuzzy granules. We can define it as follows: for  $\forall p \in P, \forall A \subseteq R$ , namely  $A = \{r_1, r_2, \dots, r_k\}, (k \leq |R|)$ , then the fuzzy granular vector of  $p$  on the attribute subset  $A$  is

$$\hat{N}_A(p) = \frac{N_{r_1}(p)}{r_1} + \frac{N_{r_2}(p)}{r_2} + \dots + \frac{N_{r_k}(p)}{r_k} = \sum_{j=1}^k \frac{N_{r_j}(p)}{r_j} \quad (4)$$



Similarly, here “+” denotes union operator and “-” represents separator. Furthermore, the module of fuzzy granular vector induced by  $p$  on attribute subset  $A$  can be denoted as:

$$|\hat{N}_A(p)| = \sum_{r \in A} |N_r(p)| \quad (5)$$

Given these definitions mentioned above, operators and metrics between fuzzy granules can also be denoted as follows. Let  $N_r(p)$  and  $N_r(v)$  be fuzzy granules of instances  $p$  and  $v$  on attribute  $r$  respectively, four operators  $\cap, \cup, \sim,$  and  $\oplus$  can be defined as follows:

$$N_r(p) \cap N_r(v) = \sum_{j=1}^m \frac{s_r(p_j, p) + s_r(p_j, v)}{1 + s_r(p_j, p) + s_r(p_j, v)} p_j \quad (6)$$

$$N_r(p) \cup N_r(v) = \sum_{j=1}^m \frac{s_r(p_j, p) s_r(p_j, v)}{1 + (1 - s_r(p_j, p))(1 - s_r(p_j, v))} p_j \quad (7)$$

$$\sim N_r(p) = \sum_{j=1}^m \frac{1 - s_r(p_j, p)}{p_j} \quad (8)$$

$$\begin{aligned} N_r(s) \oplus N_r(t) &= \sum_{j=1}^m \frac{\max\{N_r(p_j) \cup N_r(p), N_r(p_j) \cup N_r(v)\}}{p_j} \\ &\quad - \sum_{j=1}^m \frac{\min\{N_r(p_j) \cap N_r(p), N_r(p_j) \cap N_r(v)\}}{p_j} \end{aligned} \quad (9)$$

Next, operators and metrics of fuzzy granular vector can be designed on the basis of them. For  $\forall p, v \in P$ , consider two fuzzy granular vectors  $\hat{N}_A(p) = \sum_{j=1}^k \frac{N_{r_j}(p)}{r_j}, \hat{N}_A(v) = \sum_{j=1}^k \frac{N_{r_j}(v)}{r_j}$  on  $A \subseteq R, r_i \in A, i = 1, 2, \dots, k$ . Operators and metrics between fuzzy granular vectors can be defined as:

$$\hat{N}_A(p) \cap \hat{N}_A(v) = \sum_{j=1}^k \frac{N_{r_j}(p) \cap N_{r_j}(v)}{r_j} \quad (10)$$

$$\hat{N}_A(p) \cup \hat{N}_A(v) = \sum_{j=1}^k \frac{N_{r_j}(p) \cup N_{r_j}(v)}{r_j} \quad (11)$$

$$\sim \hat{N}_A(p) = \sum_{j=1}^k \frac{\sim N_{r_j}(p)}{r_j} \quad (12)$$

$$\hat{N}_A(p) \oplus \hat{N}_A(v) = \sum_{j=1}^k \frac{N_{r_j}(p) \oplus N_{r_j}(v)}{r_j} \quad (13)$$

Here, “+” represents union; “-” denotes separator. Based on these definitions, the distance between fuzzy granular vectors can be described as

$$\hat{s}_A(\hat{N}_A(p), \hat{N}_A(v)) = \frac{1}{|A| * |P|} \sum_{r \in A} \frac{|N_r(p) \oplus N_r(v)|}{|N_r(p) \cup N_r(v)|} \quad (14)$$

Below we give two theorems about the distance and the monotony of fuzzy granular vectors.

TABLE 2. Fuzzy Granulation instance Set.

$P R$	$r_1$	$r_2$	$r_3$
$p_1$	0.5	0.2	0.2
$p_2$	0.1	0.3	0.3
$p_3$	0.1	0.2	0.4
$p_4$	0.3	0.3	0.1

Theorem 1: For  $\forall p, v \in P$ , the distance of fuzzy granular vector satisfies to the following property:

$$0 \leq \hat{s}(\hat{N}_R(p), \hat{N}_R(v)) \leq 1 \quad (15)$$

Proof: Equation (2) illustrates  $N_r(p) = \sum_{i=1}^t \frac{s_r(p, p_i)}{p_i}, N_r(v) = \sum_{i=1}^t \frac{s_r(v, p_i)}{p_i}$ , and Equation (1) demonstrates  $s_r(p, p_i), s_r(v, p_i) \in [0, 1]$ . Therefore,  $|N_r(p)| = \sum_{u \in P} s_r(p, u) \in [0, |P|]$  is given by Equation (3). We have  $\hat{N}_R(p) = \sum_{j=1}^t \frac{N_{r_j}(p)}{r_j(p)}$  and  $\hat{N}_R(v) = \sum_{j=1}^t \frac{N_{r_j}(v)}{r_j(v)}$  by Equation (4). Further,  $\forall r \in R, 0 \leq \frac{|N_r(p) \oplus N_r(v)|}{|N_r(p) \cup N_r(v)|} \leq |P|$  can be given by Equation (5), (7) and (9). So Equation  $0 \leq \sum_{r \in R} \frac{|N_r(p) \oplus N_r(v)|}{|N_r(p) \cup N_r(v)|} \leq |R| * |P|$  is established. Divide both sides of the inequality by  $|R| * |P|$  to get  $\hat{s}(\hat{N}_R(p), \hat{N}_R(v)) = \frac{1}{|R| * |P|} \sum_{r \in R} \frac{|N_r(p) \oplus N_r(v)|}{|N_r(p) \cup N_r(v)|}$ , that is,  $0 \leq \hat{s}(\hat{N}_R(p), \hat{N}_R(v)) \leq 1$  is established.  $\square$

Theorem 2 (Monotony): For  $\forall p \in P$ , the attribute subset  $A$  and  $F$  satisfy  $A \subseteq F \subseteq R$ , and let  $\hat{N}_A(p)$  and  $\hat{N}_F(p)$  be fuzzy granular vectors of  $p$  on  $A$  and  $F$  respectively. Then  $|\hat{N}_A(p)| \leq |\hat{N}_F(p)|$  is established.

Proof: Equation (4) gives  $\hat{N}_A(p) = \sum_{j=1}^k \frac{N_{r_j}(p)}{r_j}$  and  $\hat{N}_F(p) = \sum_{j=1}^g \frac{N_{r_j}(p)}{r_j}$ , where  $g \leq k$ . For  $\forall r \in A$ , its fuzzy granule is  $N_r(p)$ . Since  $A \subseteq F, r \in F$ . Its fuzzy granule satisfies  $N_r(p) \in \hat{N}_F(p), |A| \leq |F|$ . So  $\sum_{r \in A} |N_r(p)| \leq \sum_{r \in F} |N_r(p)|$ , that is,  $|\hat{N}_A(p)| \leq |\hat{N}_F(p)|$  is established.  $\square$

The distance of fuzzy granular vector is regarded as the metric of similarity between fuzzy granular vectors. The smaller distance, the more similar fuzzy granular vectors are. Otherwise, the bigger distance, the less similar fuzzy granular vectors are. We have now give an example.

Example 1: As shown in Table 2, given a instance set  $P = \{p_1, p_2, p_3, p_4\}$  and an attribute set  $R = \{r_1, r_2, r_3\}$ , granulation process is as follows.

We take the instance  $p_1$  as the example. If fuzzy granulation is performed according to the attribute  $r_1$ , the fuzzy granule of  $p_1$  can be calculated as follows:  $s_{r_1}(p_1, p_1) = |0.5 - 0.5| = 0, s_{r_1}(p_1, p_2) = |0.5 - 0.1| = 0.4, s_{r_1}(p_1, p_3) = |0.5 - 0.1| = 0.4, s_{r_1}(p_1, p_4) = |0.5 - 0.3| = 0.2$ , since  $N_{r_1}(p_1) = s_{r_1}(p_1, p_1)/p_1 + s_{r_1}(p_1, p_2)/p_2 + s_{r_1}(p_1, p_3)/p_3 + s_{r_1}(p_1, p_4)/p_4$ , the fuzzy granule of  $p_1$  on  $r_1$  is  $N_{r_1}(p_1) = 0/p_1 + 0.4/p_2 + 0.4/p_3 + 0.2/p_4$ . Its cardinal is  $|N_{r_1}(p_1)| = 0 + 0.4 + 0.4 + 0.2 = 1$ . Similarly, the fuzzy granule of  $p_1$  on  $r_2$  is  $N_{r_2}(p_1) = 0/p_1 + 0.1/p_2 + 0/p_3 + 0.1/p_4$ , of which cardinal is  $|N_{r_2}(p_1)| = 0.2$ . The fuzzy granule of  $p_1$  on  $r_3$  is denoted as  $N_{r_3}(p_1) = 0/p_1 + 0.1/p_2 + 0.2/p_3 + 0.1/p_4$  and its cardinal is represented as  $|N_{r_3}(p_1)| = 0.4$ . After

that, the fuzzy granular vector of  $p_1$  on  $R$  can be defined as  $\hat{N}_R(p_1) = \sum_{j=1}^3 \frac{N_{r_j}(p_1)}{r_j}$  and its module can be calculated as  $|\hat{N}_R(p_1)| = |N_{r_1}(p_1)| + |N_{r_2}(p_1)| + |N_{r_3}(p_1)| = 1 + 0.2 + 0.4 = 1.6$ . Other instances listed in Table 2 are also granulated by the similar calculation.

### B. LOCAL TOPOLOGY OF FUZZY GRANULAR VECTOR

Let  $L_{p_i}$  be a  $(k + 1) \times (k + 1)$  matrix denoting local topology of fuzzy granular vector induced by  $p_i$ . For example, one element of  $L_{p_i}$ ,  $L_{p_i}(a, b) = \hat{s}_R(\hat{N}_R(x_a), \hat{N}_R(x_b))$  denotes the distance between fuzzy granular vector of  $x_a$  and fuzzy granular vector of  $x_b$  on attribute  $R$ , where  $x_1 = p_i, \{x_2, x_3, \dots, x_{k+1}\}$  are  $p_i$ 's  $k$  nearest neighbors of fuzzy granular vectors. Similarly,  $L_{q_j}$  is also a  $(k + 1) \times (k + 1)$  matrix denoting local topology of fuzzy granular vector induced by  $q_j$ . Let  $M$  be a  $m \times n$  matrix, where  $M^{ij}$  is the distance between  $L_{p_i}$  and  $L_{q_j}$ . As we know, there are  $k!$  permutations in the  $k$  nearest neighbors of the fuzzy granular vector from  $p_i$ . In other words, there are  $k!$  variants in  $L_{p_i}$ . Let  $\{L_{p_i}\}_h$  represent its  $h^{th}$  variant. Similarly,  $L_{q_j}$  has also  $k!$  variants and  $\{L_{q_j}\}_h$  denotes its  $h^{th}$  variant. For  $\forall M^{ij} \in M_{m \times n}$ , the calculation of  $M^{ij}$  is equivalent to calculating the distance between  $L_{p_i}$  and  $L_{q_j}$ ,  $dist(L_{p_i}, L_{q_j})$ , which can be given according to Equation  $M^{ij} = \exp(-dist(L_{p_i}, L_{q_j})/\sigma^2)$ . Here,  $dist(L_{p_i}, L_{q_j}) = \min_{1 \leq h \leq k!} \{\min(dis_1(h), dis_2(h))\}$ . In this Equation,  $dis_1(h) = \|\{L_{q_j}\}_h - \pi_1 L_{p_i}\|_F$ , and  $dis_2(h) = \|L_{p_i} - \pi_2 \{L_{q_j}\}_h\|_F$ , where  $\pi_1 = trace(L_{p_i}^T \{L_{q_j}\}_h) / trace(L_{p_i}^T L_{p_i})$ ,  $\pi_2 = trace(\{L_{q_j}\}_h^T L_{p_i}) / trace(\{L_{q_j}\}_h^T \{L_{q_j}\}_h)$  ( $\|\cdot\|_F$  represents Frobenius norm). We give the theorem and proof as follows.

**Theorem 3:** Given two  $(k + 1) \times (k + 1)$  distance matrices  $L_1$  and  $L_2$ ,  $\pi_2 = trace(L_2^T L_1) / trace(L_2^T L_2)$  can be got by minimizing  $\|L_1 - \pi_2 L_2\|_F$ . Symmetrically,  $\pi_1 = trace(L_1^T L_2) / trace(L_1^T L_1)$  can also be got by minimizing  $\|L_2 - \pi_1 L_1\|_F$ .

*Proof:* The aim is to find  $\pi_2$  to minimize  $\|L_1 - \pi_2 L_2\|_F$ . Here  $\|\cdot\|_F$  expresses Frobenius norm. This is equivalent to solving  $\pi_2 = argmin_{\pi_2} \|L_1 - \pi_2 L_2\|_F$ . It is easy to verify  $\|L_1 - \pi_2 L_2\|_F = trace(L_1^T L_1) - 2\pi_2 trace(L_1^T L_2) + \pi_2^2 trace(L_2^T L_2)$  according to Frobenius norm. Since  $trace(L_1^T L_1)$  is constant, the problem is equivalent to finding the solution on  $\pi_2 = argmin_{\pi_2} \{\pi_2^2 trace(L_2^T L_2) - 2\pi_2 trace(L_1^T L_2)\}$ . Differentiating with respect to  $\pi_2$ , this Equation implies  $2\pi_2 trace(L_2^T L_2) = 2trace(L_1^T L_2)$ . Further, it implies  $\pi_2 = trace(L_1^T L_2) / trace(L_2^T L_2)$ . Similarly, we can also find  $\pi_1 = trace(L_1^T L_2) / trace(L_1^T L_1)$ .  $\square$

To calculate matrix  $M$ , the comparison of all matrices  $L$  need to be calculated. When computing  $L_{p_i}$  and  $L_{q_j}$ , we assume that  $p_i$  and  $q_j$  are matched. However, it is unknown that  $k$  nearest neighbors of fuzzy granular vector of  $p_i$  matches that of  $q_j$ . To find the optimum matching,  $k!$  permutations need to be considered. It is easy to do, because local pattern and  $k$  are very small positive integer. As demonstrated in Theorem 3, we describe how to rescale optimally to match each other. In other words, we need to consider all possible

matching between two local patterns. That is to compute  $dist(L_{p_i}, L_{q_j})$  and return the distance calculated from the best possible match.

### C. MANIFOLD ALIGNMENT WITHOUT PREDETERMINING CORRESPONDENCE

In this section, to find the best matching between  $P$  and  $Q$ , a loss function  $LS(\alpha, \beta)$  is constructed by fusing their local topology. By solving the  $LS(\alpha, \beta)$ , the best mappings,  $\alpha, \beta$ , can be found. This computing process can be equal to solving the problem of generalized eigenvalue. We first give some notions as follows.

Let  $D_p$  be a diagonal matrix, where  $D_p^{ii} = \sum_j M_p^{ij}$  and  $V_p = D_p - M_p$ .  $M_q^{ij}$  denotes the distance between  $q_i$  and  $q_j$ , and  $D_q$  is also a diagonal matrix, where  $D_q^{ii} = \sum_j M_q^{ij}$ ,  $V_q = D_q - M_q$ . Let  $Z = \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix}$  and  $D = \begin{pmatrix} D_p & 0 \\ 0 & D_q \end{pmatrix}$ . We use matrix multiplication to map the vector  $p_i$  into a scalar, that is  $\alpha^T p_i$  ( $\alpha$  is a  $t \times 1$  matrix). Similarly, we have  $\beta^T q_j$  to obtain a scalar ( $\beta$  is a  $s \times 1$  matrix). Let  $\xi = (\alpha^T, \beta^T)^T$ .  $LS(\alpha, \beta)$  is a loss function, where  $\mu_0$  is weight of the first term of  $LS(\alpha, \beta)$ .  $\Psi_1$  is a  $m \times m$  diagonal matrix. Here  $\Psi_1^{ii} = \sum_j M^{ij}$ .  $\Psi_2$  is a  $m \times n$  matrix, where  $\Psi_2^{ij} = M^{ij}$ .  $\Psi_3$  is a  $n \times m$  matrix, where  $\Psi_3^{ji} = M^{ij}$ .  $\Psi_4$  is a  $n \times n$  matrix. Here  $\Psi_4^{ii} = \sum_j M^{ji}$ . We have  $V = \begin{pmatrix} V_p + \mu \Psi_1 & -\mu \Psi_2 \\ -\mu \Psi_3 & V_q + \mu \Psi_4 \end{pmatrix}$ . The aim is to minimize the loss function

$$LS(\alpha, \beta) = \mu \sum_{i,j} (\alpha^T p_i - \beta^T q_j)^2 M^{ij} + \mu_1 \sum_{i,j} (\alpha^T p_i - \alpha^T p_j)^2 M_p^{ij} + \mu_2 \sum_{i,j} (\beta^T q_i - \beta^T q_j)^2 M_q^{ij}$$

Here weights satisfy to  $\mu_1 + \mu_2 = 1$ . The first term of  $LS(\alpha, \beta)$  is to penalize the difference between  $P$  and  $Q$  on the matched local topology in fuzzy granular space. If  $M^{ij}$  is small, it can be inferred that  $L_{p_i}$  and  $L_{q_j}$  are close in the new space. Otherwise, if  $M^{ij}$  is big, we may conclude  $L_{p_i}$  is far distant from  $L_{q_j}$  in the new space. The second and third terms imply similarity of local topology inside  $P$  and  $Q$ . The optimal solution of this loss function  $LS(\alpha, \beta)$  is provided in Theorem 4.

**Theorem 4:** Solving the loss function  $LS(\alpha, \beta)$  is equivalent to calculating the minimum eigenvectors of the generalized eigenvalue decomposition  $ZVZ^T \xi = \lambda ZDZ^T \xi$ , which can optimal mappings to align  $P$  and  $Q$ .

*Proof:* The aim is to assure  $LS(\alpha, \beta) = \xi^T ZVZ^T \xi$  is established. Two manifolds can be aligned and their embedding structure can be discovered by joining two graphs using matrix  $V$ . To eliminate arbitrary proportionality scaling in the embedding,  $\alpha^T P D_p P^T \alpha + \beta^T Q D_q Q^T \beta = \xi^T ZDZ^T \xi = 1$  need to be as an constraint condition to be added. The matrices  $D_p$  and  $D_q$  give a measure on the vertices of the graph, where vertices can be seen as

TABLE 3. FGMA algorithm.

Input:	instance sets $P$ and $Q$ , parameter $K$ of nearest neighbor
Output:	The corresponding between $P$ and $Q$
1	Normalize the attribute values of instances to the interval $[0, 1]$ .
2	For each instance $p \in P$ , $q \in Q$ , run Step 3 to 6.
3	For $p \in P$ on the atom attribute $r_j \in R$ , fuzzily granulate instance to obtain fuzzy granule $N_{r_j}(p)$ .
4	Form fuzzy granular vector of $p$ , $\hat{N}_R(p) = \sum_{j=1}^{ R } \frac{N_{r_j}(p)}{r_j}$ .
5	For $q \in Q$ on the atom attribute $e_j \in E$ , fuzzily granulate instance to obtain fuzzy granule $N_{e_j}(q)$ .
6	Construct fuzzy granular vector of $q$ , $\hat{N}_E(q) = \sum_{j=1}^{ E } \frac{N_{e_j}(q)}{e_j}$ .
7	Build a connection on the local topology $M^{ij} = e^{-dist(L_{p_i}, L_{q_j})/\sigma^2}$ , where $L_{p_i}$ , $L_{q_j}$ and $M$ are defined in Section III.
8	Connect manifold $\mathcal{P}$ and $\mathcal{Q}$ , then compute matrix $V$ , $Z$ and $D$ . They represent model of joint structure.
9	Calculate the best mappings to implement dimensionality reduction of joint structure. The $d$ dimensional projection is obtained by eigenvectors $\xi_1, \xi_2, \dots, \xi_d$ w.r.t. $d$ minimum eigenvalues of the generalized eigenvalues decomposition $ZVZ^T\xi = \lambda ZDZ^T\xi$ .
10	Find the corresponding relationship between $P$ and $Q$ : $\alpha$ is the top $t$ rows of $[\xi_1, \xi_2, \dots, \xi_d]$ . $\beta$ is the next $s$ rows of $[\xi_1, \xi_2, \dots, \xi_d]$ . For any $i$ and $j$ , $\alpha^T p_i$ and $\beta^T q_j$ have the same dimensionality in the new space such that they can directly compare.

instances. If the value  $D_p^{ii}$  or  $D_q^{ii}$  is small, it implies  $p_i$  or  $q_i$  is less essential. If there were not the constraint condition, all instances projected into new space might be indistinguishable. Finally, the optimum problem can be converted into solving the Equation  $\text{argmin}_{\xi: \xi^T ZDZ^T \xi = 1} \{LS(\alpha, \beta)\} = \text{argmin}_{\xi: \xi^T ZDZ^T \xi = 1} \{\xi^T ZVZ^T \xi\}$ . Build Lagrange equation  $Lg(\xi, \lambda) = \xi^T ZVZ^T \xi - \lambda(\xi^T ZDZ^T \xi - 1)$ , then we have the partial derivative of the function  $Lg(\xi, \lambda)$  on  $\xi$  and can obtain  $ZVZ^T \xi = \lambda ZDZ^T \xi$ . The solution of the Equation  $LS(\alpha, \beta)$  is equal to solving the minimum eigenvectors of the generalized eigenvalue regarding  $ZVZ^T \xi = \lambda ZDZ^T \xi$ . Suppose that  $\xi_1, \xi_2, \dots, \xi_d$  are the minimum eigenvectors with respect to the  $d$  lowest eigenvalues of the Equation  $ZVZ^T \xi = \lambda ZDZ^T \xi$  and  $\alpha$  consists of the top  $t$  rows of matrix  $[\xi_1, \xi_2, \dots, \xi_d]$  and  $\beta$  is composed of the next  $s$  rows of matrix  $[\xi_1, \xi_2, \dots, \xi_d]$ . Then,  $\alpha^T p_i$  and  $\beta^T q_j$  have the same dimensionality and can be compared directly.  $\square$

The algorithm is mainly divided into three phase, instance granulation, local topology construction, and eigenvector solving, which are shown in Table 3.

## VI. EXPERIMENTAL ANALYSIS

In this section, we adopted recall and precision to evaluate the performance of algorithm. Here  $recall = \frac{TP}{TP+FN}$ ,  $precision = \frac{TP}{TP+FP}$ , where  $TP$  represents true positive,  $FN$  denotes false negative, and  $FP$  is false positive.

We employed Oxford 5K image dataset and Alzheimer's disease voice dataset to verify performance of the algorithm. Oxford 5K is a widely used in Oxford public standard dataset. It consists of 5062 landmark building images collected from Flickr. There are 11 categories, and each category has 5 query images, and total is 55 query images. The dataset contains many types of image changes, such as translation, rotation, and perspective transformation. For all datasets we adopted the standard evaluation protocols and reported Average Precision and Average Recall. we follow standard practice and resize the images to multiple and extract crops of  $256 \times 256$  pixels. In this paper, we used Histogram of Oriented Gradients (HOG) [56] and Speeded Up Robust

Features (SURF) [57] as image features. In image processing cases,  $P$  and  $Q$  represent image. We need to match point pairs in  $P$  and  $Q$ .  $p_i$  and  $q_j$  are point in  $P$  and  $Q$  respectively.  $R$ ,  $A$ , and  $E$  denote feature (attribute) set.  $r_i$ ,  $a_j$ , and  $e_k$  express feature (attribute) value respectively. We can describe HOG feature via computing and counting gradient direction histogram from local area inside image. Since the shape and appearance of local area can be well embodied via gradient direction, we adopted HOG feature descriptor to characterize Local details around the point of image. The image was split into several blocks. The center of each block with the size of  $16 \times 16$  was the focus point. Then each block was evenly divided into 4 cells and amplitudes and directions of these cells were computed. The gradient direction was divided into 8 directions on average, and then accumulate the gradient amplitude of the pixels in each cell in the same gradient direction value range to obtain an 8-dimensional cell feature vector. Connected them together to construct a 32-dimensional vector, which is used as the descriptor of the focus point.

The SURF is widely used target feature extraction and matching algorithm. It has efficient and stable characteristics and can be employed in scenes that require real-time operations such as target recognition and tracking locking. The basic idea of the SURF algorithm is derived from the scale-invariant feature transform (SIFT) algorithm [58], but it adopts a fast approximation method in specific search, neighbor feature description, and descriptor matching, which makes the execution efficiency and stability better than the SIFT algorithm. For example, in terms of feature point search, the SURF algorithm uses a square filter instead of the Gaussian filter adopted in the SIFT algorithm, and with the help of the concept of integral image, the convolution operation of image and the Gaussian differential template are converted into addition and subtraction operations on the integral image. Therefore, its computing speed is improved. HAAR features and integrated images are used to decrease computational complexity. SURF is similar to SIFT. Here, 64-dimensional feature vector was obtained by calculating horizontal and

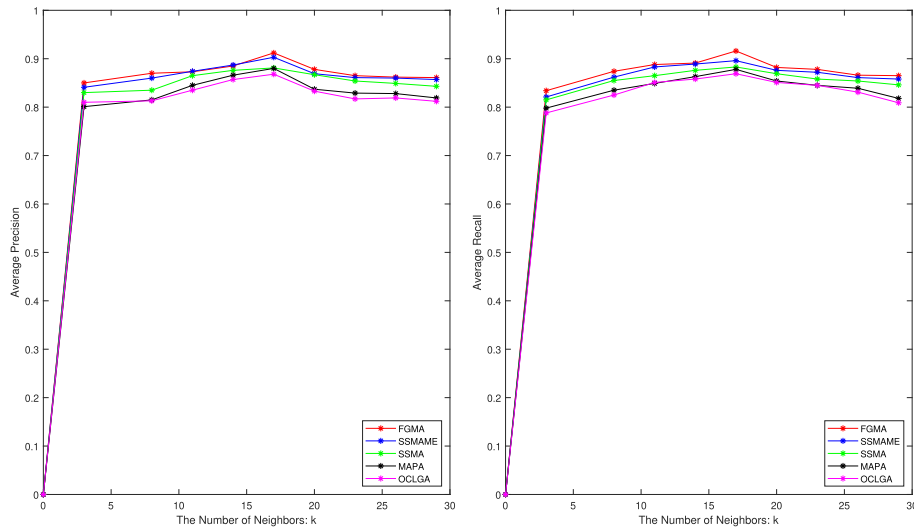


FIGURE 2. Results of Image Fuzzy Granule Manifold Alignment.

vertical HAAR wavelet features in a certain area of feature points. Figure 2 compares the performance of FGMA, semi-supervised manifold alignment (SSMA) [59], manifold alignment using Procrustes analysis (MAPA) [60], semi-supervised manifold alignment with multi-graph embedding (SSMAME) [61], and only considering local geometric alignment (OCLGA) methods. The performance refers to average precision and average recall in the experiments.

As shown in the left side of Figure 2, the average precision is shown. Its abscissa represents the number of local neighbor of fuzzy granular vector, and the ordinate denotes the average precision. When  $k = 17$ , FGMA reached a peak of 0.912, and SSMAME, SSMA, MAPA, and OCLGA achieved 0.903, 0.881, 0.880, and 0.868 respectively. Average precision increased by 1.00%, 3.52%, 3.64%, and 5.07% respectively. The average precision curve of the three methods shows a shape with a high middle and low sides. This shows that the performance is better when the number of local neighbor of fuzzy granular vector is between 10 to 20. Whether  $k$  is too large or too small, it may lead to be on the decline for performance. Overall, FGMA outperforms other four methods. The graph on the right of Figure 2 shows the average recall rate. From the curve shape, FGMA envelopes OCLGA, and OCLGA envelopes SSMA. This illustrated that in terms of average recall rate, FGMA performed OCLGA and OCLGA was superior to SSMA. From the peak of the curve, when  $k = 17$ , FGMA got the maximum value of 0.916. SSMAME, SSMA, MAPA, and OCLGA reached their maximum values of 0.896, 0.883, 0.878, and 0.869 respectively (i.e., 2.23%, 3.74%, 4.33%, and 5.41% improvement respectively). Similarly, the curve of average recall also demonstrates a shape with a high middle and low sides. Whether  $k$  is too large or too small, performance may decline. If parameter  $k$  is too large, much more noise will be imported so that performance may also fall; otherwise, if parameter  $k$  is too small,

local topology will be not embedded in matching process as constraint condition. Figure 3 shows the result of using the fuzzy granule manifold alignment, where HOG and SURF were employed to extract raw features of images.

Besides the image dataset, we also adopted voice data to verify performance of the algorithm. The Alzheimer's disease dataset came from the University of Pittsburgh, and the data was stored in form of speech and text from participants including elderly controls, people with possible Alzheimer's Disease, and people with other dementia diagnoses. There are 1263 instances in the corpus. Here,  $P$  and  $Q$  represent two voices. Point pairs in  $P$  and  $Q$  need to be compared.  $p_i$  and  $q_j$  denote point in  $P$  and  $Q$  respectively.  $R$ ,  $A$ , and  $E$  denote feature (MFCC and GFCC) set.  $r_i$ ,  $a_j$ , and  $e_k$  are feature value respectively. We know that Mel Cepstral Coefficients (MFCC) and Gammatone Cepstral Coefficients (GFCC) are highly robust, and the two methods are used as features to extract speech. First, the voice instances were down-converted to 22050 Hz and normalized. A window with a frame length of 512 (about 23ms) and a frame shift of 256 were adopted to divide the sound instance into multiple frames. Next, extract MFCC feature and GFCC feature for each voice frame separately. For MFCC, we selected the first 15 dimensions, its first-order difference, and its second-order difference, which were concatenated to obtain 42-dimensional features. For the GFCC, we selected the first 12-dimensional features, its first-order difference, and second-order difference, which were connected to get 33-dimensional features. Deleted blank frames and scrambled the dataset. As shown in Table 4, when the number of neighbors  $k = 7$ , the precision of FGMA reached maximum of 0.921, while SSMAME, SSMA, MAPA, and OCLGA got 0.905, 0.887, 0.879, and 0.868 (i.e., improvement by 1.77%, 3.83%, 4.78%, and 6.11% respectively). Next, we gave commutaion time because improvement of



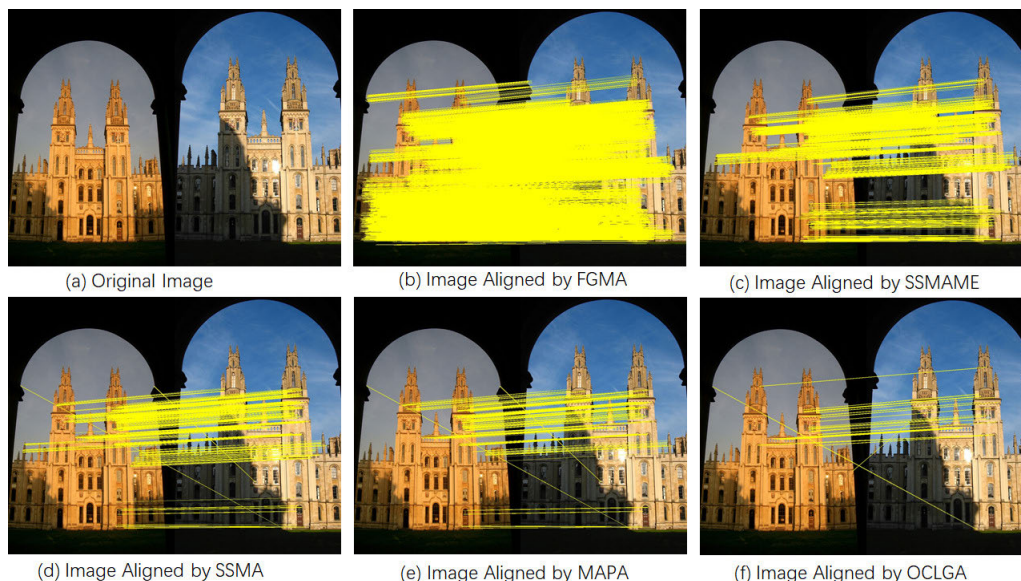


FIGURE 3. Alignment of key points between images using different algorithms.

TABLE 4. Alignment of voice collection manifolds.

Methods	The number of Neighbors	Average Precision	Average Recall	Computation Time (seconds)
FGMA	7	0.921	0.933	2.58
SSMAME	7	0.905	0.912	2.67
SSMA	7	0.887	0.879	2.52
MAPA	7	0.879	0.873	2.49
OCLGA	7	0.868	0.871	2.38

efficiency is also depended on running time. As demonstrated in Table 4, compared with the four algorithms, FGMA is better than SSMAME (i.e., improvement by 3.37%); OCLGA, MAPA, and SSMA outperformed FGMA and improved by 7.75%, 3.49%, 2.33%, respectively. In terms of recall rate, compared with SSMAME, SSMA, MAPA, and OCLGA, FGMA increased by 2.30%, 6.14%, 6.87%, and 7.12% respectively. In addition, Figure 2 also illustrated how average precision and average recall changed with the number of neighbors in the image dataset.

In summary, whether dataset is image or voice, FGMA is slightly superior to SSMAME, and SSMAME performs better than SSMA, MAPA, and OCLGA. The reasons are as follows: on the one hand, FGMA considers the global characteristics from the perspective of fuzzy granule; on the other hand, it not only considers the local topology of fuzzy granular vector induced by instance point, but also establishes an objective function to obtain an optimal solution.

### VII. DISCUSSION

In the paper, we propose a novel fuzzy granule manifold alignment method without predetermining correspondence. The approach can find correspondence between different fuzzy granular spaces, making it possible to align between instance points denoted via different features. In addition to

giving a theoretical analysis, we also apply it to the real world of image feature point alignment and Alzheimer’s speech instance matching. In future work, we will consider local and distributed granulation to further enhance algorithm efficiency.

### REFERENCES

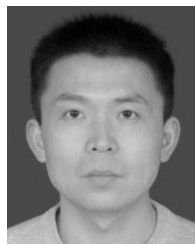
- [1] C. Bregler and S. M. Omohundro, “Nonlinear manifold learning for visual speech recognition,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Cambridge, MA, USA, Jun. 1995, pp. 494–499.
- [2] C. Bregler and M. S. Omohundro, “Nonlinear image interpolation using manifold learning,” in *Proc. Adv. Neural Inf. Process. Syst.*, Denver, CO, USA, 1994, pp. 973–980.
- [3] V. D. Silva and B. Joshua Tenenbaum, “Global versus local methods in nonlinear dimensionality reduction,” in *Proc. 16th Annu. Neural Inf. Process. Syst. Conf. (NIPS)*, Vancouver, BC, Canada, vol. 15, 2003, pp. 721–728.
- [4] H. S. Seung, “Cognition: The manifold ways of perception,” *Science*, vol. 290, no. 5500, pp. 2268–2269, Dec. 2000.
- [5] J. B. Tenenbaum, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.
- [6] S. T. Roweis, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
- [7] I. Soo Lim, P. de Heras Ciechowski, S. Sarni, and D. Thalmann, “Planar arrangement of high-dimensional biomedical data sets by isomap coordinates,” in *Proc. 16th IEEE Symp. Comput.-Based Med. Syst.*, Jun. 2003, pp. 50–55.
- [8] R. Verma, P. Khurd, and C. Davatzikos, “On analyzing diffusion tensor images by identifying manifold structure using isomaps,” *IEEE Trans. Med. Imag.*, vol. 26, no. 6, pp. 772–778, Jun. 2007.

- [9] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 14, 2002, pp. 585–591.
- [10] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [11] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 16, no. 153, 2004, pp. 153–160.
- [12] W. Yu, X. Teng, and C. Liu, "Face recognition using discriminant locality preserving projections," *Image Vis. Comput.*, vol. 24, no. 3, pp. 239–248, Mar. 2006.
- [13] J. Lu and Y.-P. Tan, "Regularized locality preserving projections and its extensions for face recognition," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 40, no. 3, pp. 958–963, Jun. 2010.
- [14] J. Gui, Z. Sun, W. Jia, R. Hu, Y. Lei, and S. Ji, "Discriminant sparse neighborhood preserving embedding for face recognition," *Pattern Recognit.*, vol. 45, no. 8, pp. 2884–2893, Aug. 2012.
- [15] C. Wang and S. Mahadevan, "Manifold alignment without correspondence," in *Proc. 21st Int. Joint Conf. Artif. Intell. (IJCAI)*, Pasadena, CA, USA, 2009, pp. 1273–1278.
- [16] E. Kokiopoulou and Y. Saad, "Orthogonal neighborhood preserving projections: A projection-based dimensionality reduction technique," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2143–2156, Dec. 2007.
- [17] H.-T. Chen, H.-W. Chang, and T.-L. Liu, "Local discriminant embedding and its variants," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 846–853.
- [18] X. Peng, J. Feng, S. Xiao, W. Yau, J. T. Zhou, and S. Yang, "Structured autoencoders for subspace clustering," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5076–5086, Oct. 2018.
- [19] X. Peng, Z. Huang, J. Lv, H. Zhu, and J. T. Zhou, "COMIC: Multi-view clustering without parameter selection," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, Jun. 2019, pp. 8925–8934.
- [20] X. Peng, H. Zhu, J. Feng, C. Shen, H. Zhang, and J. T. Zhou, "Deep clustering with sample-assignment invariance prior," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Dec. 31, 2019, doi: 10.1109/TNNLS.2019.2958324.
- [21] L. A. Zadeh, "Fuzzy sets and information granularity," in *Advances in Fuzzy Set Theory and Applications*, M. Gupta, R. Ragade, and R. Yager, Eds. Amsterdam, The Netherlands: North-Holland, vol. 2, 1979, pp. 3–18.
- [22] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets Syst.*, vol. 90, no. 2, pp. 111–127, Sep. 1997.
- [23] J. R. Hobbs, "Granularity," in *Proc. IJCAI*, Los Angeles, CA, USA, 1985, pp. 432–435.
- [24] T. Y. Lin, "Granular computing on binary relations II: Rough set representations and belief functions," *Rough Sets and Knowledge Discovery*, vol. 1. Berlin, Germany: Physica-Verlag, 1998, pp. 121–140.
- [25] R. R. Yager and D. Filev, "Operations for granular computing: Mixing words and numbers," in *Proc. IEEE Int. Conf. Fuzzy Syst. IEEE World Congr. Comput. Intell.*, May 1998, pp. 123–128.
- [26] Y. Y. Yao, "Granular computing: Basic issues and possible solutions," in *Proc. Joint Conf. Inf. Sci.*, Atlantic City, NJ, USA: Duke University/Association for Intelligent Machinery, 2000, pp. 186–189.
- [27] Y. Y. Yao, "Information granulation and rough set approximation," *Int. J. Intell. Syst.*, vol. 16, no. 1, p. 87–104, 2001.
- [28] Y. Y. Yao, "A partition model of granular computing," *Transactions on Rough Sets I (Lecture Notes in Computer Science)*, vol. 3100. Berlin, Germany: Springer, 2004, pp. 232–253.
- [29] Y. Yao, "Granular computing: Past, present and future," in *Proc. IEEE Int. Conf. Granular Comput.*, Hangzhou, China, Aug. 2008, pp. 80–85.
- [30] Y. Yao, "Interpreting concept learning in cognitive informatics and granular computing," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 39, no. 4, pp. 855–866, Aug. 2009.
- [31] W. Pedrycz, A. Skowron, and V. Kreinovich, *Handbook of Granular Computing*, 1st ed. Hoboken, NJ, USA: Wiley, 2008.
- [32] L. Zhang and B. Zhang, *Quotient Space Based Problem Solving: A Theoretical Foundation of Granular Computing*. Beijing, China: Tsinghua Univ. Press, 2014.
- [33] D. Miao, D. Li, and Y. Yao, *Uncertainty and Granular Computing*. Beijing, China: Science Press (in Chinese), 2014.
- [34] G.-Y. Wang, Q.-H. Zhang, X.-A. Ma, and Q.-S. Yang, "Granular computing models for knowledge uncertainty," *J. Softw.*, vol. 22, no. 4, pp. 676–694, Jun. 2011.
- [35] J. Liang and Y. Qian, "Information granularity and entropy theory in information system," (in Chinese), *Sci. China E, Inf. Sci.*, vol. 38, no. 12, pp. 2048–2065, 2008.
- [36] J. T. Yao, A. V. Vasilakos, and W. Pedrycz, "Granular computing: Perspectives and challenges," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1977–1989, Dec. 2013.
- [37] W. Wu, C. Gao, and T. Li, "Ordered Granular Labeled Structures and Rough Approximations," *J. Comput. Res. Develop.*, vol. 51, no. 12, pp. 2623–2632, 2014.
- [38] Y. Qian, H. Zhang, F. Li, Q. Hu, and J. Liang, "Set-based granular computing: A lattice model," *Int. J. Approx. Reasoning*, vol. 55, no. 3, pp. 834–852, Mar. 2014.
- [39] Q. Hu, Z. Xie, and D. Yu, "Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation," *Pattern Recognit.*, vol. 40, no. 12, pp. 3509–3521, Dec. 2007.
- [40] W. Xu, J. Pang, and S. Luo, "A novel cognitive system model and approach to transformation of information granules," *Int. J. Approx. Reasoning*, vol. 55, no. 3, pp. 853–866, Mar. 2014.
- [41] J. P. Herbert and J. Yao, "A granular computing framework for self-organizing maps," *Neurocomputing*, vol. 72, nos. 13–15, pp. 2865–2872, Aug. 2009.
- [42] A. Jankowski, A. Skowron, and R. W. Swiniarski, "Perspectives on uncertainty and risk in rough sets and interactive rough-granular computing," *Fundamenta Informaticae*, vol. 129, nos. 1–2, pp. 69–84, 2014.
- [43] D. Dubois and H. Prade, "Bridging gaps between several forms of granular computing," *Granular Comput.*, vol. 1, no. 2, pp. 115–126, Jun. 2016.
- [44] G. Peters and R. Weber, "DCC: A framework for dynamic granular clustering," *Granular Comput.*, vol. 1, no. 1, pp. 1–11, Mar. 2016.
- [45] J. Li, C. Mei, W. Xu, and Y. Qian, "Concept learning via granular computing: A cognitive viewpoint," *Inf. Sci.*, vol. 298, pp. 447–467, Mar. 2015.
- [46] W. Xu and W. Li, "Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 366–379, Feb. 2016.
- [47] G. Chiaselotti, D. Ciucci, and T. Gentile, "Simple graphs in granular computing," *Inf. Sci.*, vols. 340–341, pp. 279–304, May 2016.
- [48] S. Salehi, A. Selamat, and H. Fujita, "Systematic mapping study on granular computing," *Knowl.-Based Syst.*, vol. 80, pp. 78–97, May 2015.
- [49] R. Al-Hmouz, W. Pedrycz, and A. Balamash, "Description and prediction of time series: A general framework of granular computing," *Expert Syst. Appl.*, vol. 42, no. 10, pp. 4830–4839, Jun. 2015.
- [50] G. Wang, J. Yang, and J. Xu, "Granular computing: From granularity optimization to multi-granularity joint problem solving," *Granular Comput.*, vol. 2, no. 3, pp. 105–120, Sep. 2017.
- [51] J. Xu, G. Wang, T. Li, and W. Pedrycz, "Local-density-based optimal granulation and manifold information granule description," *IEEE Trans. Cybern.*, vol. 48, no. 10, pp. 2795–2808, Oct. 2018.
- [52] X. Li, J. Zhou, and W. Pedrycz, "Linking granular computing, big data and decision making: A case study in urban path planning," *Soft Comput.*, vol. 24, no. 10, pp. 7435–7450, May 2020.
- [53] W. Li, Y. Chen, and Y. Song, "Boosted K-nearest neighbor classifiers based on fuzzy granules," *Knowl.-Based Syst.*, vol. 195, pp. 1–13, Feb. 2020.
- [54] W. Li, Y. Chen, H. Hu, and C. Tang, "Using granule to search privacy preserving voice in home IoT systems," *IEEE Access*, vol. 8, pp. 31957–31969, 2020.
- [55] C. Mencar and W. Pedrycz, "Granular counting of uncertain data," *Fuzzy Sets Syst.*, vol. 387, pp. 108–126, May 2020.
- [56] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Diego, CA, USA, Jun. 2005, pp. 886–893.
- [57] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Computer Vision—ECCV*. Berlin, Germany: Springer, 2006, pp. 404–417.
- [58] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Kerkyra, Greece, vol. 2, pp. 1150–1157, Sep. 1999, doi: 10.1109/ICCV.1999.790410.
- [59] J. Wang, X. Zhang, X. Li, and J. Du, "Semi-supervised manifold alignment with few correspondences," *Neurocomputing*, vol. 230, pp. 322–331, Mar. 2017.
- [60] C. Wang and S. Mahadevan, "Manifold alignment using procrustes analysis," in *Proc. 25th Int. Conf. Mach. Learn. (ICML)*, vol. 230, 2008, pp. 1120–1127.
- [61] C.-B. Huang, T. A. Abeo, X.-Z. Luo, X.-J. Shen, J.-P. Gou, and D.-J. Niu, "Semi-supervised manifold alignment with multi-graph embedding," *Multimedia Tools Appl.*, vol. 79, nos. 27–28, pp. 20241–20262, Jul. 2020.



**WEI LI** received the Ph.D. degree from Xiamen University, China, in 2013. He was a Visiting Scholar with the Department of Computer Science, University of Massachusetts, Boston, MA, USA, from June 2018 to June 2019. He is currently an Associate Professor with the School of Computer and Information Engineering, Xiamen University of Technology, China. He has published over 20 articles in journals and conferences in these areas. His research interests include machine

learning, fuzzy modeling and granular computing, big data, and information security. He is a Committee Member of the China Computer Federation.



**CHAO TANG** was born in Hefei, Anhui, China. He received the M.S. degree from Shanxi University, Taiyuan, China, in 2009, and the Ph.D. degree in artificial intelligence from Xiamen University, Xiamen, China, in 2014. He is currently an Associate Professor with the School of Artificial Intelligence and Big Data, Hefei University, China. His project work and research interests include machine learning, computer vision, and human-action recognition.



**JIANWU XUE** received the bachelor's degree in electronic information engineering from the Xiamen University of Technology, China, in 2017, where he is currently pursuing the master's degree. His main research interests include artificial intelligence and computer graphics.



**YUMIN CHEN** (Member, IEEE) received the M.E. degree in computer application technology from Nanchang University, China, in 2005, and the Ph.D. degree in pattern recognition and intelligence system from Tongji University, China, in 2010. From November 2018 to November 2019, he was a Visiting Scholar supported by the Education Department of Fujian Province with the Department of Mathematics and Statistics, Mississippi State University, USA. He is currently

a Professor with the College of Computer and Information Engineering, Xiamen University of Technology. He has published over 50 articles in refereed journals and conferences. His research interests include machine learning, rough sets, and granular computing.



**QIANG ZHANG** received the B.S. degree from the Wuhan University of Science and Technology, in 2014. He is currently pursuing the master's degree with the School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, China. His research interests include artificial intelligence, computer vision, and granular computing.



**XUEBAI ZHANG** received the B.S. degree from Xiamen University, China, in 2011, the M.S. degree in computer science and information management from Providence University, Taiwan, in 2013, and the Ph.D. degree from the Institute of Computer Science and Engineering, National Chiao Tung University, Taiwan, in 2018. Since 2019, she has been an Assistant Professor with the School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, China.

She has authored over ten peer-reviewed international journals and conferences. Her research interests include human-computer interaction, eye tracking technique, game-based learning, digital learning, artificial intelligence, and the Internet of Thing.



**YIFANG GAO** is currently pursuing the master's degree with the School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, China. His research interests include artificial intelligence, data mining, and machine learning.

...