Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 16, No. 2, pp. 195–212, 2020

A NOVEL IMAGE RETRIEVAL APPROACH BASED ON SEMANTIC AFFINITY*

Juxiang Zhou, Xiaodong Liu, Jianhou Gan and Tianwei Xu

Abstract: Feature expression and similarity metric are two crucial steps for content-based image retrieval. In this paper, an novel image retrieval approach based on semantic affinity is proposed through exploiting discriminative features and capturing the latent data manifold. Different from the traditional practice, firstly Axiomatic Fuzzy sets (AFS) is introduced to masterly embed the semantic into original image features for forming a new semantic feature space. Then different semantic descriptions are exploited for different images according real image attribute distribution, which effectively reflect the semantic difference between images. Finally, a semantic affinity metric is defined for image matching based on the similarities of semantic descriptions. It is noteworthy that this proposed metric consider not only the similarities between two images but also their their respective neighbors, which is totally different from the traditional distance metrics model only base on pair-wise distance of images. With accurate description and reflection of semantic affinity, the proposed method can further enhance the stability and robustness of similarity relationship of images and thereby leads to improvement of the retrieval performance significantly. Extensive experiments on benchmark databases verify that it can achieve a good performance in completing the face retrieval task based on pixel-level grayscale features and the natural image retrieval task based on CNNs depth features.

Key words: *content based image retrieval, semantic metric, axiomatic·fuzzy xet*

Mathematics Subject Classification: *68U10*

1 Introduction

Content Based Image Retrieval (CBIR) was proposed in 1990s [7] to overcome the limitations of text-based image retrieval, which is suffered from the demanding time and labor requirements of annotation [26]. CBIR has been an active research topic in the fields of computer vision for decades [20]. The process of CBIR is to retrieve a ranked set of images that are relevant to a query image in terms of similarity. In a CBIR task, image features description and similarity metric directly affect the retrieval performance.

For extracting image features, many traditional CBIR approaches use hand-crafted low-level features that describe the image content such as color, texture and shape [4, 10–12, 21, 28–30, 34]. Because of the insurmountable semantic gap between the low-level representation of images and their higher-level concepts, more recent studies have focused on seeking for image representation with richer semantic. Many existing effective methods adopt bag of words (BoW) models [3, 33], vector locally aggregated descriptors (VLAD) [6] or Fisher

*This work is supported by some NFSC projects (Nos.61673082, 61462097, 61602321), and Application Infrastructure Projects of Science and Technology Plan in Yunnan Province (No.2016FD022), and Yunnan Key Laboratory of Smart Education.

vectors (FV) to encode the low-level visual features such as SIFT or SURF. With the successful applications in image recognition, convolutional neural networks (CNNs) have been introduced to perform the CBIR task as high-level features extractor [2, 20, 26, 33]. Compared with hand-crafted features, CNNs can achieve superior performance resulting from the capability of learning automatically complex features through training multiple layers of convolutional filters [20]. However, as a learning-based feature, CNNs features are extracted with the deep learning model trained for classification task in a data-driven manner. As a result, especially when the amount of data is small, or the data is from a particular field of images which are entirely different from the images used for training the pre-trained model, the learned features may not well reflect the visual content characteristics and semantic information of retrieval images [35]. Compared with the outstanding performance in the field of image recognition, the ability of learning-based features still needs to be improved in applications of image retrieval.

As another decisive factor for better retrieval performance, distance metric is used to compare the similarity between the query image and each image in the database. To achieve more accurate retrieval, the CBIR system should employ an effective similarity matching which can accurately characterize and quantify the perceptual similarities [1]. The most common used distance metrics for similarity matching in CBIR domain include Euclidean, CityBlock and so on. Despite the success of utilizing the common distance metrics in literatures, there are also some obvious limitations. First, they ignore the neighboring feature bins. Second, the neighborhood context of image is not considered, and the internal structure of the image data cannot be reflected. Finally, semantic similarities between images are not exploited and used. Ideally, the similarity between images should reflect the relevance in semantics, but it is difficult due to the intrinsic "semantic gap" problem [35]. Thus, it is still one of the challenging issues to find an adequate and robust distance metric in the field of CBIR [1].

Considering the above challenges, we need to embed semantics into extracted features and employ the context affinity to similarity matching. Motivated by this idea, Axiomatic Fuzzy Set (AFS) theory is introduced to define a fuzzy based semantic affinity metric for CBIR in this paper, which is specifically inspired by unique advantage with strong semantic representation capability as well as exploiting the information on the latent data distribution. Recently, AFS theory has been applied in data clustering and recognition [8, 9, 19, 22, 24, 32]. These studies have two points in common: semantic information can be integrated into raw data features by using AFS theory, and the similarities between data are measured by semantic metric. It can be confirmed that AFS-based methods are capable of capturing subtle similarity information distributed over semantic feature subspaces to reveal the latent data distribution more accurately, and thereby lead to improved performance in data clustering.

However, unlike data clustering and recognition, CBIR has no training and learning. As natural color images include complex content and are often represented with high-dimension features, a more accurate and robust similarity measure is required. Therefore, in this paper, AFS is introduced to deal with CBIR tasks including face retrieval with raw low-level features and natural image retrieval with high-level CNNs features. Based on AFS framework, the extracted features are represented with semantic attributes to form new features space. Furthermore, discriminative semantic description on different features are exploited according to the distributions of original data and the semantics of the fuzzy sets. Then, in image matching, instead of only considering the pairwise similarity or distance between two images, the nearest neighbors of every image are employed to generate the distance affinity matrix. More specifically, we define a semantic metric to compute the semantic similarity between two image sets, where we not only consider their respective exploited semantic descriptions, but also their respective nearest neighbors. In summary, the contributions of this paper are stated as follows:

1. AFS theory is applied in CBIR to complete image retrieval task at the first time. The semantic information is subtly integrated into original image feature space to form a new semantic space, in which different semantic descriptions are exploited for different images.

2. A novel fuzzy-based semantic similarity metric is proposed for image matching. Instead of only considering the pairwise similarity or distance in the existing methods, the neighbor relationship is employed to generate the distance affinity matrix for capturing the latent and stable data structure.

3. Extensive experiments have demonstrated that the proposed method is superior to both traditional distance metrics and existing AFS-based semantic metrics.

The rest of this paper is organized as follows. In Section 2, we provide a brief introduction of AFS theory and introduce some related works. The proposed image retrieval method based semantic affinity is presented in Section 3. In Section 4, experimental results and performance comparison are given. Section 5 concludes the paper.

2 Preliminaries and Related work

2.1 AFS Theory

AFS theory was originally proposed in [14] and then extensively developed in [16–18], etc. Due to its unique characteristic in semantic representation, AFS theory has been applied in various research fields over data mining, knowledge discovery, business intelligence, financial analysis, computer vision and so on. In this section, we will give a brief introduction of AFS theory about essential notions and foundations that will be helpful to understand the related work and our study. More details can be seen in [15, 36].

As the basis of semantic representation in AFS theory, the concept is used to construct AFS fuzzy sets through logical operation, which is determined by membership functions according to distributions of the original data. The essence of this process is to imitate human's learning behavior for knowledge representation and reasoning. The AFS algebra delivers a new approach related to the semantic interpretation of fuzzy attribute. It is worth emphasizing that the information can be extracted from the AFS space rather than the original data space. Two key steps of AFS are conducting the semantic representation for feature description and computing the coherence membership functions of fuzzy sets.

2.1.1 Semantic Representation for Feature Description

Let X be a set of people, in which each sample has several features including age, health, degree of education and treasure. Let $M = \{m_1, m_2, \ldots, m_n\}$ be the set of concepts on samples X, where m_1(young), m_2(about 50 years old), m_3(poor health), m_4(high education), m_5(rich) and so on, that are associated to each feature for describing the people. Using M, all possible concepts on X can be flexibly formulated through conjunction and disjunction operation of simple concepts to construct complete fuzzy sets. Specifically, for any concept set $A \subseteq M$, some complex concepts Υ can be generated by Eq.(2.1):

$$\Upsilon = \sum_{i\in I}(\prod_{m\in A_i} m), A_i \subseteq M, i \in I \tag{2.1}$$

Let M be a non-empty set. The set EM^* is defined in Eq.(2.2)

$$EM^* = \left\{ \sum_{i\in I}(\prod_{m\in A_i} m)\big|A_i \subseteq M, i \in I, I\ is\ a\ no\, empty\ indexing\ set. \right\} \tag{2.2}$$

EM^* is a set of the sums that represent all semantic combinations of the simple concepts in M.

$\boxed{2.1.2}$ Coherence Membership Functions of Fuzzy Sets

Let X be a data set and M be a set of fuzzy terms on X. For $A \subseteq M, x \in X$, we define

$$A^{\succeq}(x) = \{y \in X | x_{\succeq m} y\ for\ any\ m \in A\} \subseteq X \tag{2.3}$$

where "$\succeq$" is a linearly ordered relation. For $m \in M$, "$x_{\succeq m} y$" implies that the degree of x belonging to m is larger than or equal to that of y. $A^{\succeq}(x)$ is the set of all elements in X whose degrees of belonging to set $\prod_{m\in A} m$ are less than or equal to that of x. $A^{\succeq}(x)$ is determined by the semantic of the fuzzy terms in A and the probability distribution of the observed dataset X. Let ν be a fuzzy term on X.

The coherence membership functions are associated with a measurement over X, which can be constructed by taking the semantics of the fuzzy terms and the probability distribution of the features into account. The following definition is first introduced.

Definition 2.1 ([16]). Let ν be a fuzzy term on X. $\rho_v : X \to R^+ = [0, +\infty)$. ρ_v is called a weight function of the fuzzy term ν if ρ_v satisfies the following conditions:

1. $\rho_v(x) = 0$, if x does not belong to ν;
2. $\rho_v(x) \geq \rho_v(y)$, if the degree of x belonging to ν is larger than or equal to that of y.

Next, For each fuzzy set $\xi \in EM$, the coherence membership functions can be computed as,

$$\mu_\xi(x) = \sup_{i\in I} \inf_{\gamma\in A_i} \frac{\sum_{\mu\in A_i^{\succeq}(x)} \rho_\gamma(u) N_u}{\sum_{\mu\in X} \rho_\gamma(u) N_u}, \forall x \in X \tag{2.4}$$

where N_u is the number of samples of u and ρ is the weight function that can be defined according to the specificity of the data and the underlying semantics of ν in practice, such as triangular, Gaussian or other type of weight functions.

It can be seen that the membership functions are based on the fuzzy logic operations expressed on the observed data and the overall space by taking both fuzziness and randomness into account.

$\boxed{2.2}$ AFS-based Metric in Clustering

With AFS framework, AFS-based methods have attracted much attentions. AFS clustering algorithm was first proposed in [18]. Many extended and improved studies have been reported later [8, 19, 23, 32]. In essence, the AFS clustering approaches are to capture the underlying data structure through fuzzy membership function, and the distances between samples are represented by membership degree. Those methods are able to establish discriminative features [8]. In AFS-based semantic feature space, the distance such as Euclidean metric for overall original feature space is replaced by these AFS-based metric using the semantic similarity or distance computed by the membership degree on fuzzy set.

Having an insight into AFS-based metric, there are two ways for measurement. For two samples $x_i, x_j \in X$, the similarity between them is defined as [19,32]:

$$s_{i,j} = min\{\mu_{\xi_{x_i} \wedge \xi_{x_j}}(x_i), \mu_{\xi_{x_i} \wedge \xi_{x_j}}(x_j)\} \tag{2.5}$$

where ξ_{x_i} and ξ_{x_j} represent the extracted descriptions for x_i and x_j respectively. That means the similarity between two samples is determined by their respective semantic membership degree belonging to the fuzzy term $\xi_{x_i} \wedge \xi_{x_j}$. Here, "$\wedge$" is the fuzzy conjunction logic operation in AFS algebra.

Or, the distance between them can be defined in [8]:

$$d_{i,j} = 1 - min\{\overline{\mu}_{\xi_{x_j}}(x_i), \overline{\mu}_{\xi_{x_i}}(x_j)\} \tag{2.6}$$

$$\overline{\mu}_{\xi_{x_i}}(x_j) = \left\{ m_k \in \xi_{x_i} | \frac{\sum_{k=1}^{N_{\xi_{x_i}}} \mu_{m_k}(x_j)}{N_{\xi_{x_i}}} \right\} \tag{2.7}$$

where m_k represents each fuzzy term belonging to ξ_{x_i}. $\overline{\mu}_{\xi_{x_i}}(x_j)$ represents the mean membership degree of x_j belonging to the description of x_i, $N_{\xi_{x_i}}$ is the number of fuzzy terms in description of x_i.

In this paper, we defined a new semantic similarity metric by introducing the local neighbor relation, which is more accurate and robust for CBIR in real applications.

3 Image retrieval based on Semantic Affinity

3.1 Semantic Description in AFS

For a set of images $X = \{x_1, x_2, \ldots, x_N\}$, let $F = \{f_1, f_2, \ldots, f_d\}$ be the extracted d-dimension feature vector for every image. Based on AFS framework, we define a set of fuzzy terms $M = \{m_{i,j} | 1 \le i \le d, 1 \le j \le 3\}$, where $m_{i,1}, m_{i,2}, m_{i,3}$ represents the fuzzy concept of "small", "medium" and "large" associated with the feature f_i respectively. As mentioned above, to further compute coherence membership functions of fuzzy sets, a weight function is adopted to obtain the weight of a sample belonging to every simple concept. For simplicity, a triangular weight functions $\rho_{m_{i1}}$,$\rho_{m_{i2}}$ and $\rho_{m_{i3}}$ are used for semantic concepts "small", "medium" and"large" respectively. They are defined as follows:

$$\rho_{m_{i1}} = \frac{f_i^{max} - f_i}{f_i^{max} - f_i^{min}} \tag{3.1}$$

$$\rho_{m_{i2}} = \begin{cases} \frac{f_i - f_i^{min}}{f_i^{avg} - f_i^{min}}, f_i^{min} \le f_i \le f_i^{avg} \\ \frac{f_i^{max} - f_i}{f_i^{max} - f_i^{avg}}, f_i^{avg} < f_i \le f_i^{max} \end{cases} \tag{3.2}$$

$$\rho_{m_{i3}} = \frac{f_i - f_i^{min}}{f_i^{max} - f_i^{min}} \tag{3.3}$$

where f_i^{max} means the max value of f_i, f_i^{min} is the min value of f_i, and $f_i^{avg} = (f_i^{max} - f_i^{min})/2$. Then the original features are expanded to a more specific and explicit semantic space with $3 * d$ dimension.

Then, the coherence membership functions using Eq.(2.4) can be used to compute the membership degree of each sample x belong to any fuzzy term, which reflects the goodness of description. Obviously, not all of the fuzzy terms are good semantic descriptions to represent x. So, a salient fuzzy subset is selected to construct the good description ξ_x of x according the goodness of every fuzzy term with a selection criteria as below:

$$\xi_x = \bigwedge_{m \in M_x^{good}} m, \; M_x^{good} = \{m \in M | \mu_m(x) \geq max\{\mu_m(x)\} - \varepsilon\} \tag{3.4}$$

where ε is a small positive number to control the error threshold, which is set empirically. Because every m in M_x^{good} is good enough to represent x, the fuzzy term ξ_x can describe x very well with accurate semantic by avoiding indistinctive features, yielding sample representation that can better express the underlying semantic structure in data [8]. Finally, we can construct a semantic description set $\xi = \{\xi_{x_1}, \xi_{x_2}, \ldots, \xi_{x_N}\}$ for all image samples in AFS framework.

3.2 Affinity Matrix based on Semantic Metric

In general, generating affinity matrix based on the feature description is the crucial step for image retrieval. A robust and accurate affinity reflection exploited from semantic underlying structure will certainly achieve good performance.

In this paper, instead of only considering the pairwise similarity or distance between two data elements in the existing AFS-based distance metrics, a new semantic metric based on affinity matrix via using local neighbor relation. If two data elements are similarity, the best semantic description for one is also better to describe the other for their nearest neighbors. In essence, the affinity constraint is changed from the similarity between two images to two image groups of the nearest neighbour. Obviously, such similarity relation is more stable and robust, and can effectively reduce the influence caused by noise outlier samples. For two samples x_i and x_j, the similarity between them is defined as:

$$s_{i,j} = \sum_{k=1}^{K} (\mu_{\xi_{x_i}}(x_j^k) + \mu_{\xi_{x_j}}(x_i^k)) \tag{3.5}$$

where ξ_{x_i} and ξ_{x_j} is the good description of x_i and x_j, x_i^k and x_j^k are the k-th nearest neighbor of x_i and x_j. After computing all pair similarity $s_{i,j}$, the affinity matrix can be obtained as $S = \{s_{i,j}, 1 \geq i \geq N, 1 \geq j \geq N\}$. The matrix S is symmetric, and the element $s_{i,j}$ represents the similarity degree between samples x_i and x_j. The larger the value is, the more similar of them.

3.3 Algorithm for Image Retrieval

Actually, based on obtained affinity matrix, the retrieval task can be easily completed according the descending sort order. However, considering the difference in importance of each dimension of features for image description, we only select few significant features from the extracted feature set $F = \{f_1, f_2, \ldots, f_d\}$ to construct the semantic description set ξ for X. We define the salience value I_x^f of each feature f for image x as the number of the set $N^f \cap N^F$. Here, N^f and N^F are all subset of X, and they are the nearest neighbors of image x with respect to feature f and F respectively. The larger the salience value, the higher the significance, which also means that the feature f is more important to describe image x. This process can be viewed as a feature selection.

Therefore, the algorithm of image retrieval based on proposed AFS-based similarity affinity metric (AFSSAM) can be detailed as Algorithm 1. The process from Step 2 to Step 10 is to select significant features F_{select}. Step 11 to Step 16 is to select salient fuzzy subset from M for every image x and construct the good description ξ_x for x. Step 17 is to compute the similarity based on semantic metric, and the affinity matrix can be obtained in Step 18.

Algorithm 1: Image retrieval based on AFSSAM

Input: X:The image data matrix with size of $N \times d$; N_k: The number of nearest neighbors to compute the saliency of features; N_s: The number of selected salient features; K: The number of nearest neighbors to compute the similarity; $\varepsilon \in [0, 1]$: The membership degree threshold.

Output: S^*: The sorted affinity matrix, in which the element of s^*_{ij} presents the j-th retrieved image of i-th query image.

Normalize every feature f of X to unit interval [0,1] by linear transformation.
for each image $x \in X$ **do**
Find the N_k nearest neighbors set N^F according to the distance in overall F feature space.
  for each feature $f \in F$ **do**
  Find the N_k nearest neighbors set N^f according to the distance in f feature space.
  Compute the number of $|N^F \bigcap N^f|$ as the salience I_x^f of feature f for image x
  end for
end for
Sort the feature set F by the salience I_x^f in descending order.
Select the top-N_s salient features in sorted F as F_{select}.
for each image $x \in X$ **do**
Generate the simple concepts m_{i1}, m_{i2} and m_{i1} for each f in F_{select}
Construct the fuzzy concepts set M on x through conjunction and disjunction operation of simple concepts
Select $M_x^{good} = \{m \in M | \mu_m(x) \geq max\{\mu_m(x)\} - \varepsilon\}$
Construct the good description $\xi_x = \bigwedge_{m \in M_x^{good}} m$
end for
Compute the similarity $S = \{s_{ij}\}$, where $s_{i,j} = \sum_{k=1}^{K}(\mu_{\xi_{x_i}}(x_j^k) + \mu_{\xi_{x_j}}(x_i^k))$
Return the affinity matrix $S^* = \{s^*_{ij}\}$ obtained by sorting each row of S according to the value of $s_{i,j}$ in descending order.

4 Experiments

4.1 Databases and Experimental Instruction

In order to investigate the performance of the proposed method from different perspectives, two face databases and two nature image databases are used in our experiments. YaleB [5] consists of 165 face images of various persons under different poses and illumination conditions. It has 15 subjects are shown in 11 different conditions. ORL has 400 images of 40 subjects with 10 grayscale images per subject, where pose, illumination, and expression are diverging. Corel1k [31] contains 1000 images that are divided into 10 categories, and 100 images per category. Corel5k contains 5000 images from 50 categories having 100 images in each category.

For simplicity, pixel-based gray features are extracted directly from original images for YaleB and ORL databases. Specifically, each image of them is down-sampled to size 16×16 and then normalized to 0-mean and 1-variance. Finally, 256-dimension vectorized feature representations can be obtained for every images on both face databases. In addition, for two nature image databases, learning-based CNNs features are used in our experiments. Usually, when passing an input image through a CNN model, the outputs from convolutional layers are feature maps, in which each element corresponds to receptive field of the input image. The activations of convolutional layers tend to catch finer grained information that may be beneficial to content-based retrieval. In experiments, the VGG-verydeep-16 model [25] is employed to extract deep CNNs features using the open-source library MatConvNet [27]. All images are resized to 224×224 and sent to the network. Then the last convolutional layers "pool5" descriptors with sum-pooling aggregation are selected as image features of 512-dimension.

PCA is used for dimensionality reduction of original features extracted above. The final retained PCA dimensions are 20 for YaleB, 30 for ORL, 12 for Corel1k and 65 for Corel5k. They are all optimal setting to acquire the best retrieval performance using Euclidean distance for similarity metric which is regard as the baseline to be compared in latter experiments. Moreover, in the proposed AFSSAM, for all database, N_k is set to the number of images in every category empirically. The membership degree threshold $\varepsilon = 0.3$ in all experiments, which has been investigated and considered as a reasonable value in most of AFS related work. The value of N_s is set by searching over the ranges of 1 to number of the respective retained PCA dimensions. The value of K is set by searching over the ranges of 1 to 10 for face databases, and 1 to 20 for two Corel databases respectively. In order to analyze the rules for setting parameters, the influence of N_s and K on retrieval results have also been investigated in later experiments.

Each image in four databases is used as query image in all experiments. The retrieval performance of methods is measured in terms of average retrieval precision (ARP), average retrieval rate (ARR), and the mean average precision (mAP). For one retrieval method, the larger values of ARP, ARR and mAP obtained, the better performance it has when certain numbers of L image returned after retrieval. In all experiments, $L = 15$ for YaleB and ORL database, and $L = 100$ for Corel1k and Corel5k unless specified in some special cases.

4.2 Experimental Results

4.2.1 Performance Comparison with other Distance Metrics

First, extensive experiments are designed to compare the performance of the proposed metric with some common used distance metrics including CityBlock(4.1), D1(4.2), Euclidean(4.3), Canberra(4.4) and Chebychev(4.5) for similarity matching in CBIR, which are defined as follows:

$$d_{i,j} = \sum_{t=1}^{d} |f_t^{x_i} - f_t^{x_j}| \tag{4.1}$$

$$d_{i,j} = \sum_{t=1}^{d} \frac{|f_t^{x_i} - f_t^{x_j}|}{1 + f_t^{x_i} + f_t^{x_j}} \tag{4.2}$$

$$d_{i,j} = \sqrt{\sum_{t=1}^{d} (f_t^{x_i} - f_t^{x_j})^2} \tag{4.3}$$

$$d_{i,j} = \sum_{t=1}^{d} \frac{|f_t^{x_i} - f_t^{x_j}|}{|f_t^{x_i} + f_t^{x_j} + \delta|} \tag{4.4}$$

$$d_{i,j} = \lim_{k\to\infty} (\sum_{t=1}^{d} |f_t^{x_i} - f_t^{x_j}|^k)^{1/k} \tag{4.5}$$

where $f_t^{x_i}$ is the t-th component of feature f for image x_i, and δ is a a small equilibrium constant.

The performance comparison on YaleB and ORL databases is listed in Table 1. Obviously, the proposed metric AFSSAM achieves the highest scores all over the evaluation index. Especially on YALE database, more than 10%, 15% and 7% of ARP, ARR and mAP increase has been achieved comparing with the best performance of traditional metrics.

Fig.1-2 show the performance comparison of ARP, ARR and mAP on Corel1k and Corel5k database. We can see that when the number of L changes from 20 to 100, the ARP and ARR obtained by the proposed method are always higher than others with increasing improvements. Although the gap of them is tiny at first, it became more and more distinct especially in term of ARP. When $L = 100$ the AFSSAM can achieve 86.59% and 55.68% on Corel1k and Corel5k respectively, which are 5% and 6% higher than Euclidean metric respectively. In general, when the number of L increase, there will be more interference brought by some outliers, and the retrieval task becomes more difficult. But in this case, the superiority of the proposed method becomes more evident. This conclusion can be also verified by the results listed in Table 2, in which mAP is used to evaluate the performance of retrieval task. Furthermore, from the above experiments it is not difficult to find that the improvement brought by the proposed method on two face databases is more remarkable than these on two Corel databases.

Table 1: The performance comparison on YaleB and ORL databases

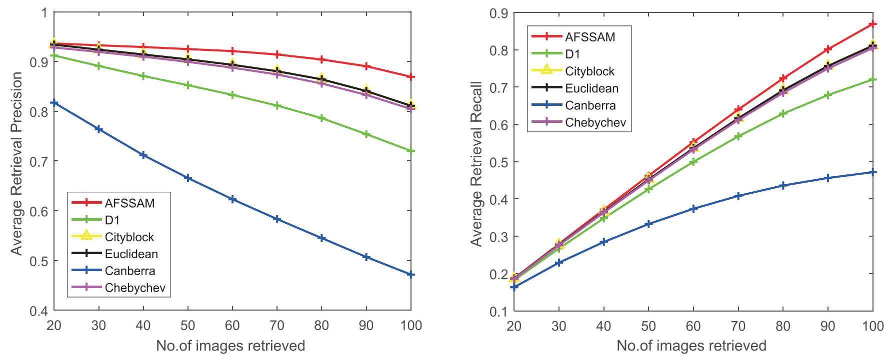
	YaleB			ORL		
	ARP	ARR	mAP	ARP	ARR	mAP
D1	0.417	0.568	0.660	0.406	0.610	0.677
CityBlock	0.518	0.706	0.751	0.486	0.729	0.754
Euclidean	0.518	0.706	0.751	0.486	0.729	0.754
Canberra	0.356	0.485	0.599	0.262	0.393	0.505
Chebychev	0.445	0.607	0.692	0.426	0.638	0.689
AFSSAM	**0.616**	**0.850**	**0.822**	**0.527**	**0.790**	**0.774**

Table 2: Performance comparison of mAP on Corel database

	D1	CityBlock	Euclidean	Canberra	Chebychev	AFSSAM
Corel1k	0.848	0.899	0.899	0.679	0.893	**0.920**
Corel5k	0.649	0.658	0.658	0.292	0.587	**0.683**

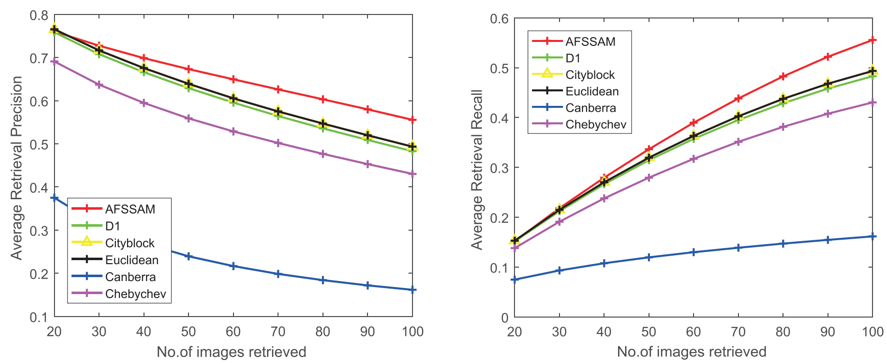
4.2.2 Performance Comparison with AFS-based Metrics

As the detailed in section 2.2, there are two common usedly AFS-based similarity or distance metrics. In order to verify the validity of the proposed semantic affinity metric, we design this experiment to compare with these two methods, as listed in Table 3 and Table 4. Here,



(a) ARP (b) ARR

Figure 1: Performance comparison with traditional distance metrics on Corel1k database



(a) ARP (b) ARR

Figure 2: Performance comparison with traditional distance metrics on Corel5k database

AFS1 and AFS2 present the metric methods using Eq.(2.5) and Eq.(2.6) respectively. And the scores of AFS1 and AFS2 are obtained by the best performance with optimal N_s.

From the results of Table 3 and Table 4, we can see that the proposed method AFSSAM performs the best on all databases and evaluation indicators. It implies that, comparing with other two AFS-based metrics, the proposed method has robust and stable superiority for affinity measurement.

From the above experimental comparison, it is convincing that the proposed method is not only competent to describe semantic features but also can capture the semantic similarity accurately. Next, we investigate the image affinity matrix constructed by the proposed method, which will intrinsically reflect how effective an affinity metric method is. Take the Corel1k as an example, the affinity matrices generated by the proposed metric, traditional Euclidean metric, AFS1, and AFS2 are given in Fig.3. It can be observed that the proposed AFSSAM produces affinity matrices with more distinct block structure and less false edges compared with other methods. This suggests the superiority of the proposed method in learning the underlying semantic data structures, potentially leading to more accurate similarity relationship. Specifically, the affinity matrix produced by AFS1 has less false edges in estimating the dissimilarities (inter-similarities) relation but the block

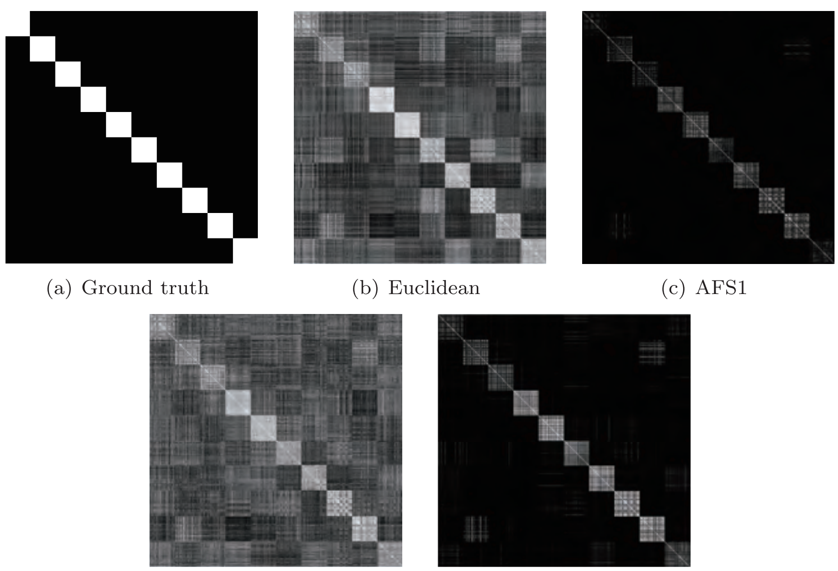
Table 3: The performance comparison with AFS-base metrics on face databases

	YaleB			ORL		
	ARP	ARR	mAP	ARP	ARR	mAP
AFS1	0.506	0.690	0.733	0.412	0.619	0.662
AFS2	0.521	0.710	0.755	0.463	0.695	0.733
AFSSAM	**0.616**	**0.840**	**0.822**	**0.527**	**0.790**	**0.774**

structure along the main diagonal for similarities (intra-similarities) is not clear and compact. Meanwhile, AFS2 and Euclidean relatively perform poor in describing the dissimilarities relation, although they produce acceptable results for measurement of similarities. The proposed metric achieves the best performance overall.

Table 4: The performance comparison with AFS-base metrics on Corel databases

	Corel1k			Corel5k		
	ARP(L=20)	APR(L=100)	mAP	ARP(L=20)	APR(L=100)	mAP
AFS1	0.925	0.822	0.897	0.677	0.447	0.589
AFS2	0.921	0.752	0.868	0.744	0.461	0.630
AFSSAM	**0.936**	**0.866**	**0.920**	**0.758**	**0.557**	**0.683**



(a) Ground truth (b) Euclidean (c) AFS1

(d) AFS2 (e) AFSSAM

Figure 3: Comparison of the affinity matrix generated by different methods on Corel1k database.

4.2.3 Performance Comparison with State-of-the-art Methods

Although this paper focuses on affinity metric, some performance comparisons with state-of-the-art methods are given in this section.Various retrieval methods using hand-crafted features or CNNs features are compared with our proposed method on Corel databases. Fig.4-5 shows some retrieval results on Corel databases in term of ARP and ARR with the number of returned images L from 20 to 100. Here, AFSSAM is the proposed method, and CNNs present the method using Euclidean distance metric with the same features as the proposed method. Other methods including "MTH" [12], "MSD" [10], "CDH" [11], "Zhou2018" [34], "prakash2012", [21], "verma2015" [28], "vipparthi2014" [30] and "vipparthi2015" [29] are all hand-crafted features, which aim to improve retrieval accuracy by integrating advantages of different visual features.

From the results shown in Fig.4, it is quite clear that the scores obtained by the proposed method AFSSAM and CNNs are significantly higher than others. It's amazing that on Corel1k, the gaps of ARP between them are always increasing from near 20% when $L = 20$ to more than 30% when $L = 100$. For ARR, such strong improvement also keeps when $L = 100$. In the same ways, on Corel5k as shown in Fig.5, the proposed method AFSSAM and CNNs methods still perform outstandingly, which achieve more than 20% improvement of ARP and ARR when $L = 100$ comparing with other methods. We have to say that CNNs features based method can obtain significant performance. Though the proposed method AFSSAM can outperform the CNNs with Euclidean distance metric.

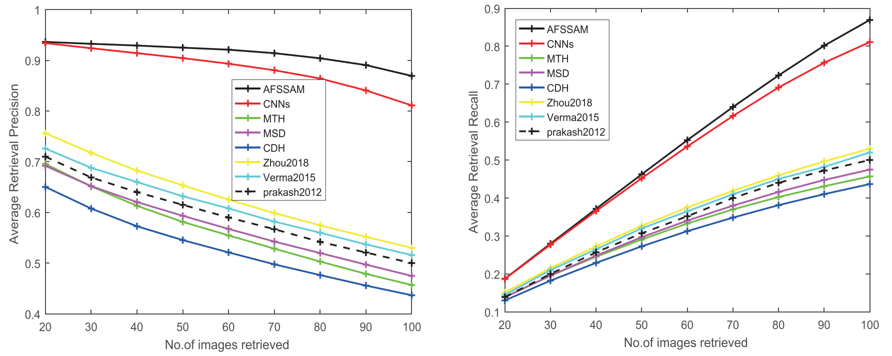
Next, a group of comparison between the proposed method and other CNNs-based methods are investigated on Corel1k database. In [13], the authors presented an effective image retrieval method by combining high-level features from convolutional neural network (CNN) model and low-level features from dot-diffused block truncation coding (DDBTC). With the fusion of the DDBTC and CNN features, a new codebook feature was generated using the two layer codebook. Table 5 lists some results reported in [13]. Even if the process of dimension reduction and similarity reweighting are employed in that methods, the proposed AFSSAM still achieves higher retrieval accuracy with only 12-dimension CNNs features after PCA reduction.

Table 5: Performance comparison with CNNs-based methods

Methods	Features	Dimension	ARP (L=20)	APR (L=100)
GL-FCF	GoogLeNet FC layer feature	1024	0.896	0.709
DL-TLCF-hier	Deep Learning Two-layer codebook features with hierarchical structure	300/1408	0.919	0.735
DL-TLCF-hier(nor)	Deep Learning Two-layer codebook features	1408	0.925	0.741
HD(DL-TLCF)	High Dimensional DL-TLCF with hierarchical structure	300/1408	0.919	0.742
HD(DL-TLCF)(nor)	High Dimensional DL-TLCF	1408	0.923	0.710
CNNs+Euclidean	VGG-Verydeep-16 Pool5 Layer	256	0.898	0.611
CNNs+AFSSAM	VGG-Verydeep-16 Pool5 Layer	12	**0.936**	**0.866**

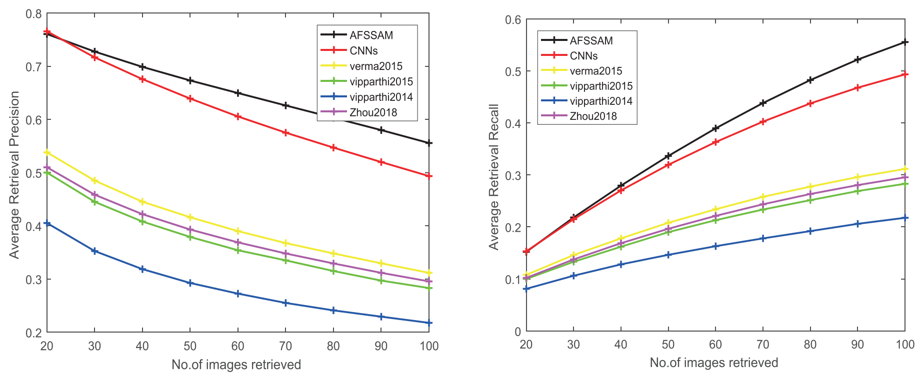
4.3 Parameter analysis

As we demonstrated in experimental setting, in our proposed method there are four parameters to be considered. N_k and ε can be easily assigned reasonable values empirically. Therefore, in this experiment we analyze the influence of different K and N_s on the results,



(a) ARP (b) ARR

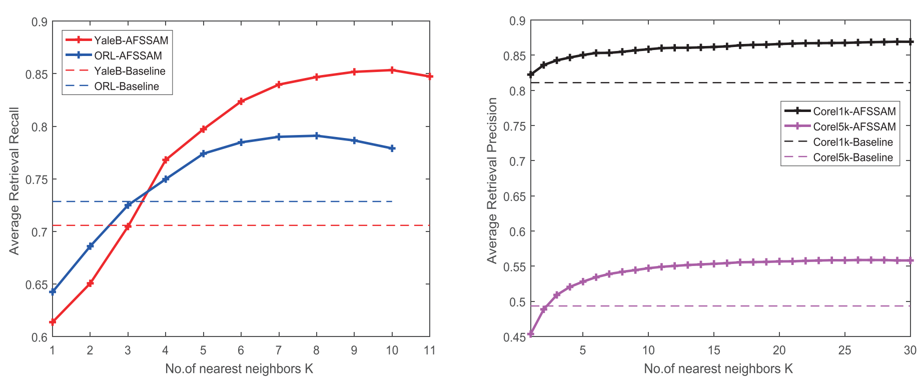
Figure 4: Performance comparison with the state of the arts methods on Corel1k database



(a) ARP (b) ARR

Figure 5: Performance comparison with the state of the arts methods on Corel5k database

by thoroughly testing a range of parameter values, as shown in Fig.6 and Fig.7. Here, the baselines for every database present the performance using the Euclidean distance.

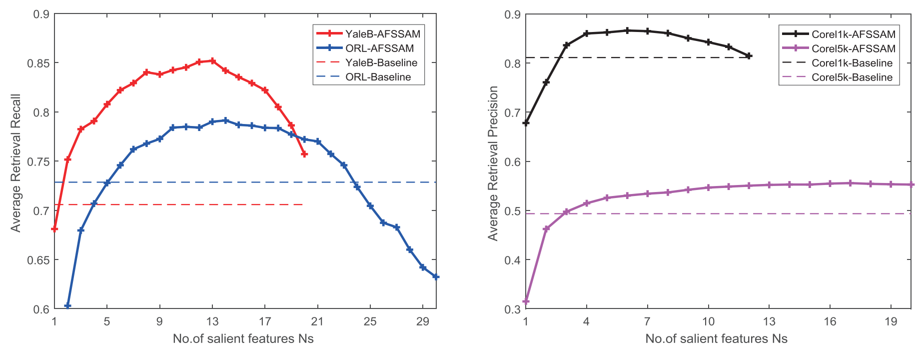


(a) YaleB and ORL databases (b) Corel1k and Corel5k databases

Figure 6: Influence of the number of nearest neighbor on the obtainable retrieval scores

According to the number of related images (YaleB 11, ORL 10, Corel1k 100, Corel5k 100) for each query images in database, the range of K is set to [1,11] and [1,30] respectively for face database and Corel database.

As can be seen, selecting a reasonable K value is important to obtain optimal retrieval results. On YaleB and ORL databases as seen in Fig.6(a), when $K \geq 4$ the performance is significantly improved in comparison to the baseline. The best scores are achieved for $K = 10$ and $K = 7$ respectively for YaleB and ORL databases. If the K gets closer to the number of the related images for query image, the performance improves slowly down or starts to fall. On two Corel databases as seen in Fig.6(b), as the parameter K increasing, the performance is also improved comparing with the baseline. When $K \geq 20$ the improvement tends to be stable.



(a) YaleB and ORL databases

(b) Corel1k and Corel5k databases

Figure 7: Influence of the number of selected salient features

Further, let us analyze the influence of N_s on the performance of the proposed method. As know N_s is the number of selected salient features to construct the description for each image. As shown in Fig.7, although the features in front position are more important to describe the image due to the concepts are sorted by the feature salience in descending order, too few feature is not enough to reflect the discriminative semantic. Along with the increase of N_s, the performance is significantly improved in comparison to the baseline. Meanwhile, the trend of improvement turns to stable or decreasing when N_s reach a certain value. specifically, the optimal retrieval results can be achieved when $N_s = 13$ for YaleB and ORL database, and $N_s = 6$ for Corel1k, and $N_s = 17$ for Corl5k.

As illustrated in experimental setting, the final feature dimensions are 20, 30, 12 and 65 for YaleB, ORL, Corel1k and Corel5k respectively, which are also the upper limit of each N_s. Thus, N_s should be in the range of a half to a quarter of feature dimension in order to achieve good results. It seems to be a promising direction of research to automatically find an optimal parameters in an efficient manner. In practice, feature extraction can be completed offline. Although the feature dimension is relatively low, the time consumed for generating affinity matrix would growth as the increasing number of images in database as well as K and N_s.

$\boxed{5}$ Conclusion

In this paper, a new affinity metric based on semantic similarity is proposed by using Axiomatic Fuzzy Set. On one hand, different from the traditional practice this method embeds the semantic concepts into original image features and exploits different semantic descriptions based on real data attribute distribution. On the other hand, in order to overcome the limitation brought by the traditional distance metrics, the similarity between not only any two images but also their respective neighbors are considered. So, the latent and stable data structure can be captured to make more accurate description and reflection of semantic affinity. Extensive experiments on four databases are used to verify the superiority of the proposed method.

Acknowledgements

We thank the anonymous reviewers.

References

[1] A. Alzu'Bi, A. Amira and N. Ramzan, Semantic content-based image retrieval: A comprehensive study, *Journal of Visual Communication & Image Representation* 32 (2015) 20–54.

[2] A. Alzu'Bi, A. Amira and N. Ramzan, Content-based image retrieval with compact deep convolutional features, *Neurocomputing* 249 (2017) 95–105.

[3] Y. Cao, C. Wang, Z. Li, L. Zhang and L. Zhang, Spatial-bag-of-features, in: *IEEEConference on Computer Vision and Pattern Recognition*, 2010, pp. 3352–3359.

[4] C. Celik and H.S. Bilge, Content based image retrieval with sparse representations andlocal feature descriptors: A comparative study, *Pattern Recognition* 68 (2017) 1–13.

[5] A.S. Georghiades, P.N. Belhumeur and D.J. Kriegman, From few to many: illumination-cone models for face recognition under variable lighting and pose, *IEEE Transactions on Pattern Analysis & Machine Intelligence* 23 (2002) 643–660.

[6] H. Jegou, M. Douze, C. Schmid and P. Perez, Aggregating local descriptors into a compact image representation, in: *Computer Vision and Pattern Recognition*, 2010, pp. 3304–3311.

[7] T. Kato, TDatabase architecture for content-based image retrieval, in: *Proceedings of SPIE- The International Society for Optical Engineering*, vol. 1662, 1992, pp. 112–123.

[8] Q. Li, Y. Ren, L. Li and W. Liu, Fuzzy based affinity learning for spectral clustering, *Pattern Recognition* 60 (2016) 531–542.

[9] Z. Li,Q. Zhang, X. Duan and Y. Wang, A novel semantic approach for multi-ethnicface recognition, *International Journal of Pattern Recognition & Artificial Intelligence* 32 (2018) 1856005.

[10] G.H. Liu, Z.Y. Li, L. Zhang and Y. Xu, Image retrieval based on micro-structure descriptor, *Pattern Recognition* 44 (2011) 2123–2133.

[11] G.H. Liu and J.Y. Yang, Content-based image retrieval using color difference histogram, *Pattern Recognition* 46 (2013) 188–198.

[12] G.H. Liu, L. Zhang, Y.K. Hou, Z.Y. Li and J.Y. Yang, Image retrieval based on multi-texton histogram, *Pattern Recognition* 43 (2010) 2380–2389.

[13] P. Liu, J.M. Guo, C.Y. Wu and D. Cai, Fusion of deep learning and compressed domain features for content based image retrieval, *IEEE Transactions on Image Processing, A Publication of the IEEE Signal Processing Society* 26 (2017) 5706–5716.

[14] X. Liu, The fuzzy theory based on afs algebras and afs structure, *Journal of Mathematical Analysis & Applications* 217 (1998) 459-478.

[15] X. Liu and W. Pedrycz, *Axiomatic Fuzzy Set Theory and Its Applications*, Studies in Fuzzi-ness & Soft Computing, vol. 244, Springer, 2009.

[16] X. Liu, W. Pedrycz, T. Chai and M. Song, The development of fuzzy rough sets with the use of structures and algebras of axiomatic fuzzy sets, *IEEE Transactions on Knowledge& Data Engineering* 21 (2009) 443–462.

[17] X. Liu, W. Pedrycz and Q. Zhang, Axiomatics fuzzy sets logic, in: *The IEEE International Conference on Fuzzy Systems*, 2003, pp. 55–60.

[18] X. Liu, W. Wang and T. Chai, The fuzzy clustering analysis based on afs theory, *IEEE Transactions on Systems Man & Cybernetics Part B* 35 (2005) 1013–1027.

[19] X. Liu, X. Wang and W. Pedrycz, Fuzzy clustering with semantic interpretation, *Applied Soft Computing* 26 (2015) 21–30.

[20] Y.H. Ng, F. Yang and L.S. Davis, Exploiting local features from deep networks for image retrieval, in: *Computer Vision and Pattern Recognition*, 2015, pp. 53–61.

[21] N. Prakash and K. Satya Prasad, HSV color motif co-occurrence matrix for content based image retrieval, *International Journal of Computer Applications* 48 (2012) 8-14.

[22] Y. Ren, Q. Li, W. Liu and L. Li, Semantic facial descriptor extraction via axiomatic fuzzy set, *Neurocomputing* 171 (2016) 1462–1474.

[23] Y. Ren, M. Song and X. Liu, New approaches to the fuzzy clustering via afs theory, *International Journal of Information and Systems Sciences* 3 (2007) 307–325.

[24] R. Sarkhel, N. Das, A.K. Saha and M. Nasipuri, A multi-objective approach towards cost effective isolated handwritten bangla character and digit recognition, *Pattern Recognition* 58 (2016) 172–189.

[25] K. Simonyan and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*, Computer Science, 2014

[26] M. Tzelepi and A. Tefas, Deep convolutional learning for content based image retrieval, *Neurocomputing* 275 (2018) 2467–2478.

[27] A. Vedaldi and K. Lenc, Matconvnet: Convolutional neural networks for matlab, 2014, pp. 689–692.

[28] M. Verma, B. Raman and S. Murala,Local extrema co-occurrence pattern for color and texture image retrieval, *Neurocomputing* 165 (2015) 255–269.

[29] S.K. Vipparthi, S. Murala and S.K. Nagar, Dual directional multi-motif xor patterns: A new feature descriptor for image indexing and retrieval, *Optik-International Journal for Light and Electron Optics* 126 (2015) 1467–1473.

[30] S.K. Vipparthi and S. Nagar, Expert image retrieval system using directional local motif xor patterns, *Expert Systems with Applications* 41 (2014) 8016–8026.

[31] J.Z. Wang and J. Li and G. Wiederhold, Simplicity: semantics-sensitive integrated matching for picture libraries, *IEEE Transactions on Pattern Analysis & Machine Intelligence* 23 (2001) 947–963.

[32] Y. Wang, X. Duan, X. Liu, C. Wang and Z. Li, A spectral clustering method with semantic interpretation based on axiomatic fuzzy set theory, *Applied Soft Computing* 64 (2018) 59–74.

[33] X.S. Wei, J.H. Luo, J. Wu and Z.H. Zhou, Selective convolutional descriptor aggregation for fine-grained image retrieval, *IEEE Transactions on Image Processing* 26 (2017) 2868–2881.

[34] J.X. Zhou, X.D. Liu, T.W. Xu, T.W.,J.H. Gan and W.Q. Liu, A new fusion approach for content based image retrieval with color histogram and local directional pattern, *International Journal of Machine Learning & Cybernetics* 9 (2018) 677–689.

[35] W. Zhou, H. Li and Q. Tian, Recent advance in content-based image retrieval: A literature survey, *pp.arXiv preprint arXiv:1706.06064*, (2017)

[36] X.D. Liu, W.J. Jia, Y.G. Wang, H.Y. Guo, Y. Ren and Z.D. Li, *Knowledge Discovery and Semantic Learning in the Framework of Axiomatic Fuzzy Set Theory*, WIREs Data Mining Knowl Discov, 2018.

Manuscript received 11 April 2018
revised 26 October 2018
accepted for publication 20 November 2018

Juxiang Zhou
School of Control Science and Engineering
Faculty of Electronic Information and Electrical Engineering
Dalian University of Technology, Dalian 116024, China
Key Laboratory of Education Informatization for Nationalities
Ministry of Education, Yunnan Normal University
Kunming, 650500, China
E-mail address: zjuxiang@126.com

Xiaodong Liu
School of Control Science and Engineering
Faculty of Electronic Information and Electrical Engineering
Dalian University of Technology, Dalian 116024, China
E-mail address: xdlious@dlut.edu.cn

Jianhou Gan
Key Laboratory of Education Informatization for Nationalities
Ministry of Education, Yunnan Normal University
Kunming, 650500, China
E-mail address: ganjh@ynnu.edu.cn

Tianwei Xu
Key Laboratory of Education Informatization for Nationalities
Ministry of Education, Yunnan Normal University
Kunming, 650500, China
E-mail address: xutianwei@ynnu.edu.cn