Image Feature Dimension Reduction Based on Improved LLE Algorithm

Long Bai
School of Physics and Electronic Engineering
Mudanjiang Normal University,
Mudanjiang 157011, Heilongjiang, China

Abstract — Locally linear embedding is an effective image feature dimension reduction algorithm, but when the sample distribution is not uniform. This algorithm has selected neighbor point improper and neighbor requires human factor k given defect. So an improved adaptive LLE algorithm is proposed. First, this algorithm calculates the sample point between Euclidean distance and cosine similarity, and then by a defined similar factor to effectively select neighbor points. At the same time, the algorithm is based on statistical analysis of the sample theory reconstruction 3σ rule to dynamically determine neighbor factor k. Experimental results show that this algorithm obtained feature after dimension reduction, and can effectively improve the accuracy of image classification, and less time-consuming, the overall performance is superior to the classical LLE algorithm.

Keywords - image feature; dimension reduction; LLE; adaptive

I. INTRODUCTION

In the field of image processing, image feature extraction is a crucial step, excellent image characteristics are beneficial to a series of subsequent processing, including classification, identification, retrieval[1-4]. Typically, the image processing more pursuit of the results accuracy, it is generally as much as possible to extract image features from different angles, including image size, type, texture, shape, etc. If all of these features are used, then the orders of magnitude up to 102 or even more, which for real-time algorithms and the system will be a huge challenge[5]. Further, the more and more quantity means good, But it will be affect the results of image processing because of the presence of redundant information and noise. Moreover, there may be a complex nonlinear relationship between high dimensional data, which is not easy to store and transfer[6]. Therefore, the research of image feature dimension reduction method has become a research hotspot in the field of image processing.

From the current situation, the application of more conventional image feature reduction methods principal include component analysis (PCA), independent component analysis (ICA), etc. the principle of these methods is simple, easy to implement, but they only apply to linear data and susceptible to the image-capturing angle influence, light and other factors. Therefore, in recent years there have been some nonlinear dimensionality reduction methods, such as: Isomap, Laplacian eigenmap, LTSA and LLE like. Where in, LLE algorithm since the principle is clear, less parameters, can be obtained global optimal solution, applicability is widely used [7]. This method considers the issue from the perspective of human cognition, by analyzing the intrinsic relationship between the data mapped from high dimension to low dimension, while maintaining geometric topology data between unchanged. However, this method still requires the presence of a high-density sampling for noise sensitive parameters need to artificially given other shortcomings. To solve these problems, some experts and scholars from different aspects of the research. Document [8] proposed a variable k nearest neighbor LLE image dimension reduction method using the sample mean and variance to dynamically select the appropriate number of neighbors, effectively improve the accuracy of image retrieval; [9] proposed a weighted LLE(WLLE) algorithm to calculate the importance of the use of thermonuclear function value for each sample point, and added to the LLE algorithm's cost function, effectively reducing the algorithm sensitive for noise; [10] for the classical LLE algorithm distance calculation methods is improved so that the sample sampling distribute more evenly , thus effectively reducing the LLE algorithm for parameter k dependence; [11] when using LLE algorithm used to calculate the geodesic distance replaces the traditional Euclidean distance while using the maps loss function to dynamically determine the number of neighbors k, so that LLE algorithm has adaptive characteristics. Unfortunately, the proposed algorithm still need to pre-k empirically given a range of options, and then dynamically select k values within this range, it can’t be regarded as completely adaptive.

Through analysis, the above document upon improved LLE algorithm, focus mainly on the change in distance calculation, while ignoring the real correlation between samples, thus reducing relatively sparse data poor dimensional effect for the distribution, in addition select the k value is not to be completely adaptive. Based on this, this paper presents a fully adaptive ALLE (Adaptive LLE, ALLE) image feature dimension reduction methods. The main innovation is that using the improved Euclidean distance and cosine similarity to characterize the degree of association between the samples, effectively improve the
uniformity of the sample and sampling of authenticity. Additionally, this paper uses sample rule to dynamically determine the number of neighbor $k$ such that the algorithm has excellent adaptability. Experimental results show that the proposed algorithm is adaptive LLE image feature dimension reduction than the classical LLE algorithm has better performance.

II. BASIC PRINCIPLES OF LLE ALGORITHM

LLE algorithm is a typical nonlinear feature dimension reduction method. Its main principle is to use local linear approximation of the overall nonlinear data, the overall data is through the transformation of the topology before and after the change, as well as local neighborhood overlap to effectively protect.

Suppose algorithm input $X = [x_1, x_2, \ldots, x_N]$ , $x_i \in R^p, i = 1, 2, \ldots, N$ the output of the algorithm $Y = [y_1, y_2, \ldots, y_N]$ and $y_i \in R^q, i = 1, 2, \ldots, N$ , $d \ll D$. The specific steps LLE algorithm is:

Step 1: find the nearest neighbor of each sample.

In the original high-dimensional space, followed by calculation of the Euclidean distance between each sample and the other samples between points, and are arranged in ascending order from the results, take the first one as a neighborhood point samples. Previously given by human.

Step 2: Calculate the partial reconstruction of the weight matrix.

Sample error function is defined by a neighbor point approximated as:

$$e(W) = \sum_{i=1}^{N} \left\| x_i - \sum_{j=1}^{N} w_{ij} y_j \right\|$$  \hspace{1cm} (1)

By using the Lagrange method, the weight matrix can be obtained by minimizing operation the function.

Among them, $w_{ij}$ is the weight for the sample $x_i$ and $y_j$ , and meet the following conditions:

$$\sum_{j=1}^{N} w_{ij} = 1$$  \hspace{1cm} (2)

$$w_{ij} = \begin{cases} 1 & \text{ if } x_i \text{ is the neighborhood of } y_j \cr 0 & \text{ otherwise} \end{cases}$$  \hspace{1cm} (3)

Step 3: Calculate the local $d$ dimensional embedding space results $Y$

After dimension reduction, high dimensional space is changed into low dimensional space, but the weight matrix of the nearest neighbor and local reconstruction remains the same. Therefore, a error function is similarly defined with the formula (1):

$$\Phi(Y) = \sum_{i=1}^{N} \left\| y_i - \sum_{j=1}^{N} w_{ij} y_j \right\|^2$$  \hspace{1cm} (4)

$y_i$ can be obtained through the formula (4) minimize operation.

Note that in the minimization process, in order to maintain the unique, should satisfy the following constraints:

$$\sum_{i=1}^{N} y_i = 1, \frac{1}{N} \sum_{i=1}^{N} y_i y_j^T = I_{d \times d}$$  \hspace{1cm} (5)

$W$ is used to represent the column $i$ of matrix $W$ , $I_i$ representing column $i$ of a unit array $N \times N$ . Then, the formula (4) can be rewritten as:

$$\Phi(Y) = \sum_{i=1}^{N} \left\| I_i - y_i y_i^T \right\| = \left\| y (1-W) \right\| = \text{tr}\{YMY^T\}$$  \hspace{1cm} (6)

Which, $M = (1-W)(1-W)^T$

Output results $Y$ is the feature vector composed, because $M$ is a sparse, symmetric and positive semi definite matrix, so by $2 - (d + 1)$ characteristic value of $M$ (in ascending order) corresponding to the feature vectors constitute the transpose of the matrix is $Y$.

III. IMPROVED ADAPTIVE ALGORITHM LLE

A. Calculation of similarity between samples

For the first problem, we believe that in the calculation of sample points $x$ and $y$ the distance, in addition to the Euclidean distance, should also consider the average distance of a sample point with all other sample points, so the distance calculation formula is:

$$d_{xy} = \frac{1}{M} \left\| x - y \right\|$$  \hspace{1cm} (7)

Among them, $\left\| x - y \right\|$ is the Euclidean distance between the two sample points, $M$ and $M_i$ respectively as the sample point , as the average distance between points and other samples. In this way you can effective improve the sampling distribution of the sample, make the structure of neighbor points more uniform and reasonable.

In addition to calculate the distance, this article should be considered to calculate the similarity between sample points, which can help identify the neighbor point and the sample point is the same class or similar category.

According to the definition of relevance \cite{12}, assuming that two random variables $X$ and $Y$ , then the correlation coefficient between them is:

$$\rho_{XY} = \frac{\text{Cov}(X,Y)}{\sqrt{D(X)} \cdot \sqrt{D(Y)}}$$  \hspace{1cm} (8)

Among them, $\text{Cov}(X,Y)$ represents the covariance between $X$ and $Y$ , $D(X)$ and $D(Y)$ respectively represents the standard deviation of $X$ and $Y$ . The larger the value of the correlation between the two variables, the
value of $\rho$ larger, when $\rho = 1$, the two variables indicates perfect correlation.

For an image, if the extracted eigenvectors have $N$, each characterize have $D$ dimension, then the similarity between the two samples is calculated as follows:

$$r_y = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{N} (y_i - \bar{y})^2}}$$

(9)

Among them, $\bar{x}$ and $\bar{y}$ are the mean value of the sample characteristic.

It can be easily seen that the distance between them as small as possible for two sample points $x_i$, but the similarity is the bigger the better, so we similarity factor is defined as follows:

$$S_y = \frac{r_y}{d_y}$$

(10)

Thus, we can calculate the similarity factor between sample points to elect a more reasonable neighbor points.

### B. dynamically selected neighborhood factors $k$

In this paper, the nearest neighbor factor $k$ selection method is mainly based on the sample distribution. Due to the calculation of the correlation between the sample points. In this paper, adopted a similarity factor as defined in (10), the formula is composed of the Euclidean distance and cosine correlation coefficient. It follows that meet or approximate Gaussian distribution. According to the law of $3\sigma$ [13], in the interval $[\mu - 3\sigma, \mu + 3\sigma]$ it will contain 99.73% of such samples. But the rule for outliers punishment too harsh, it may cause the value is too large, so this article will be adjust range of interval $[\mu, \mu + \sigma]$. Although the range is smaller, but the sample still contains up to 90%, for the partial reconstruction of the sample has a higher rationality. Specific implementation method:

For feature vector set, $X = [x_1, x_2, \cdots, x_N]$, $x_i \in \mathbb{R}^D$, $i = 1, 2, \cdots, N$, sample point $x_i$ and other features similar factors $[S_{i1}, S_{i2}, \cdots, S_{i(i-1)}, S_{i(i+1)}, \cdots, S_{i(k-1)}, S_{ik}]$ are respectively calculated, and then calculate the mean and variance of similarity factor:

$$\mu_i = \frac{1}{N-1} \sum_{j=1}^{N} S_{ij}$$

(11)

$$\sigma^2_i = \frac{1}{N-1} \sum_{j=1}^{N} (S_{ij} - \mu_i)^2$$

(12)

Finally, the number of statistical similarity factors in the interval $[\mu_i, \mu_i + \sigma_i]$ is $k$.

In summary, the proposed adaptive LLE algorithm in this article, based on the classical LLE algorithm to replace the traditional Euclidean distance factor similar step 1 with formula (10) at the same time determining factor neighbor according to the method described above, the other steps in algorithm remain unchanged.

### IV. EXPERIMENTAL ANALYSIS

In order to objectively evaluate the effective performance of the adaptive LLE algorithm proposed in this paper designed the experiment: First deal image feature dimension reduction with algorithm, and then use SVM algorithm image classification.

In this paper, simulation experiments using commonly two sets of image data: ORL and JAFFE, which contains different angles, different expressions of face images, very suitable for image to verify the performance of processing algorithms, particularly described in Table I. Experimental environment: individual PC, dual-core processor, clocked at 2.66GHz, RAM 3GB, software version Matlab7.11.

<table>
<thead>
<tr>
<th>Table I. Dataset description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dataset</td>
</tr>
<tr>
<td>ORL</td>
</tr>
<tr>
<td>JAFFE</td>
</tr>
</tbody>
</table>

From Table I, ORL and JAFFE data sets contains many sample characteristics, if all of these features used to achieve image classification, the computation complexity of the algorithm is too high, so the first for feature dimensionality reduction. To make the results more convincing, algorithms and classical article LLE algorithm [11] Improved LLE algorithm were compared. Comparison of the major classification accuracy of index, which is calculated as follows:

$$Acc = \frac{NC}{NA} \times 100\%$$

(13)

Where, $NC$ is the samples that classified correctly, $NA$ represents the total samples.

Usually case, the LLE algorithm, when sample data are densely distributed, the size of dimensionality reduction factor neighbors $k$ little effect, but the distribution is relatively sparse samples, in order to be effective, should take a larger number of neighborhood factors, so run classical LLE algorithm, this paper use value of $k$ are 10,30,50, when running the literature [11] algorithm, set $k$ value in the range of 10 to 50, while the adaptive algorithm is given in this paper. Under the same experimental conditions, a total of 20 repeated experiments were conducted. The specific classification result Fruit shown in Figure 1 and Figure 2.
We can see from Figures 1 and 2 category results, adaptive LLE proposed algorithm in this article is much better than the classical LLE algorithm and [11] algorithm. Specifically, since the sample distribution is more sparse, so classical LLE algorithm increases the value of $k$, classification effect becomes better, average classification accuracy were: 75.96%, 79.45%, 83.42% (ORL), 68.85%, 71.96%, 77.54% (JAFFE), the overall relatively not high. Algorithm with the literature [11] algorithm similar, respectively: 91.67%, 89.90% (ORL), 82.96%, 81.04% (JAFFE), and sometimes even a single recognition accuracy rate is even lower than the literature [11] algorithm, but the overall performance is better. This is mainly because the point of this article when you select neighbors, using Euclidean distance and cosine similarity combining strategy makes the selected neighbor points more scientific and rational, dimension reduction better, therefore classification accuracy to be higher.

Below these three algorithms used at the time were compared, still use of the above experimental data and methods, in order to avoid a single experiment appears occasionally deviation, respectively, each algorithm we have carried out the experiment repeat10 times, each time from a data set of 10 randomly selected images, embedded timing function in Matlab source code. Finally, the average time of the algorithm used, the specific results shown in table II.

### TABLE II. COMPARE OF ALGORITHM TIME CONSUMING (S)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>k=10</th>
<th>k=30</th>
<th>k=50</th>
<th>Ref.[11]</th>
<th>This paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORL</td>
<td>34.7</td>
<td>106.2</td>
<td>175.5</td>
<td>163.4</td>
<td>127.1</td>
</tr>
<tr>
<td>JAFFE</td>
<td>42.9</td>
<td>117.8</td>
<td>190.6</td>
<td>182.1</td>
<td>135.6</td>
</tr>
</tbody>
</table>

Careful observation can be seen in Table 2, along with the added value of $k$ classical LLE algorithm used time is gradually increasing, the literature [11] algorithm uses time slightly lower than the classical LLE algorithm 50, and the algorithm uses time significantly lower than the literature [11] algorithm, equal to the classical LLE algorithm take quite 30. This is mainly because the literature [11] algorithm in the selection of the value of $k$ consuming too much time, and the algorithm used time mainly when consumed in the similarity factor calculation, but still significantly better than the literature [11] algorithm.

Comprehensive comparison, although the algorithm on the classification accuracy rate is slightly higher than the literature [11] algorithm, but the algorithm uses less time, so the overall performance better, more suitable for practical engineering application.

V. CONCLUSION

This paper focuses on LLE algorithm in image feature dimension reduction application. For classical LLE algorithm neighbor points selected too mechanical and neighbor requires human factor given defect, it proposes a combination of Euclidean distance and cosine similarity neighbor point selection method to effectively prevent the local reconstruction failures from sparse samples. Meanwhile this paper, a rule based on 2 neighborhood factors dynamic adaptive selection strategy, so that the algorithm adaptable, to avoid the influence of human experience dimensionality reduction results. Experimental result shows that the improved adaptive LLE proposed algorithm has good performance, it is possible to maintain a high level of classification accuracy under the circumstance. The algorithm used to get less time, is very suitable for image feature dimension reduction applications.

ACKNOWLEDGMENT

This work was supported by Mudanjiang Normal University provincial key innovative pre-research project (SY201218).

REFERENCES


