ISSN:0975-9646

Sanjay Kumar et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014, 6449-6452

A Survey on Image Feature Selection Techniques

Dr. Sanjay Kumar , Er. Ankur Chauhan

Jaipur National University

Abstract—**Image is a work of art that describes or store visual perception. That has an exactly same appearance to subject – normally a physical object or a person. Therefore, images providing a representation of real time physical objects. Additionally, representation of 2 dimensional data and their patterns analysis is basic interest of the given paper. As a part of object recognition, the image and their objects can be recognised using their pattern. But there are some issues are involved for finding the essential patterns from image data. First, the amount of data to be analysed, Secondly, a simple and single class may have a huge amount of data to be process due to their class definition. In this context, dimensionality reduction techniques are helpful for reducing amount of data and finding the accurate pattern form the image data. Therefore, in this paper PCA, LDA and ICA methods are studied. These techniques are the most popular and frequently used method of pattern extraction from different kind of data types i.e. image and huge datasets. In addition of that, these techniques are helpful in classifying the objects based on their extracted pattern. The presented paper introduces discussion about these techniques.**

Keywords— ***PCA, LDA, ICA, Pattern extraction, feature selection.***

I. INTRODUCTION

Image is a basically 2D data for representing a real time object. But to store the objects image contains a large amount of data. If an image is compared with another image by pixel values, than a large amount of resources are consumed. Therefore, Data Patter analysis can be played an important role for recognizing objects in a given image. Pattern analysis is an essential application of the data mining. In this context training is performed in order to discover meaningful information from the available data. Additionally, to identify the patterns form data feature selection approaches are applied. The presented paper involves the study of feature selection techniques i.e. LDA, PCA and ICA. These techniques are statistically analyse data to recover meaningful patterns (correlation and difference) from image data in order to recognize hidden pattern form similar class data.

In object recognition process a single class can have more than one class definitions; therefore an object can be defined by the multiple views. So, a technique is required to recover the hidden patterns from the data using the extracted features. As given in figure 1, the learning process required extracting the features form the images, these feature vectors are correlated using the feature selection process, to reduce the amount of data dimensionality reduction techniques are helpful.

Therefore in the next section some frequently used techniques are described.

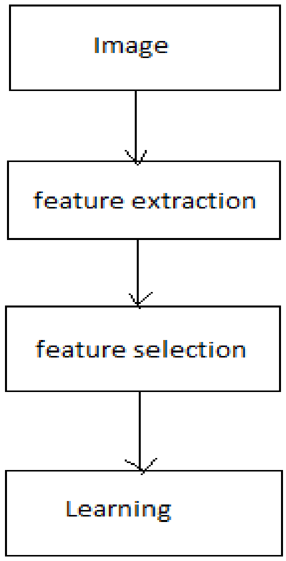

Figure 1 object recognition

II. FEATURE SELECTION TECHNIQUES

This section includes a brief introduction about the feature selection techniques that are frequently used now in these days.

A. Independent Component Analysis

ICA is a general-purpose statistical technique in which perceived random data are linearly transformed into components that are extremely independent from each other, and concurrently have “interesting” distributions. ICA can be formulated as the estimation of a latent variable model. The in-built notion of maximum non-Gaussianity can be used to develop different objective functions whose optimization supports the approximation of the ICA model. Otherwise, one may use more conventional notions like maximum likelihood approximation or minimization of reciprocal information to estimate ICA; somewhat unexpectedly, these methods are about equivalent. Computationally, a very efficient technique Fast-ICA algorithm is also available for actual calculation. Applications of ICA can be found in many different areas such as biomedical signal processing, audio processing, image processing, and telecommunications [11].

B. Principal Component Analysis

Principal Components Analysis is a way of recognizing patterns in data, and expressing the data in such a manner as to focus their differences and similarities. Subsequently patterns in data may be complex to discover in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analysing data.

The key advantage of PCA is that once we have found the patterns in the data, and you compress the data, i.e. by reducing the number of dimensions, without much loss of information. This technique used in image compression [1].

C. Linear Discriminant Analysis

There are many possible techniques for classification of data. Principal Component Analysis and Linear Discriminant Analysis are commonly used techniques for dimensionality reduction and data classification. Linear Discriminant Analysis easily handles the case where the within-class frequencies are unequal and their performances have been examined on randomly generated test data. This technique maximizes the proportion of between-class variance to the within-class variance in any specific data set in that way promising maximal separability. The Linear Discriminant Analysis is used for classification issues such as speech recognition. The key difference among LDA and PCA is that PCA perform feature classification and LDA works for data classification. The shape and location of the inventive data sets changes when transformed to a different space in PCA, on the other hand LDA doesn't change the location but only attempts to offer more class separability and induce a decision region among the given classes. This technique also supports to better recognize the distribution of the feature data [2].

III. ALGORITHM STUDY

In this section of the presented paper we learn about the different feature selection algorithm namely PCA, LDA and ICA.

A. Principal Component Analysis (PCA)

It is an traditional method of face recognition which is based on the Karhunen-Loeve Transform (KLT), works on dimensionality reduction in face recognition. Turk and Pentland used PCA exclusively for face recognition[4]. PCA computes a set of subspace basis vectors for a database of face images. These basis vectors are representation of an images which is correspond to a face – like structures named Eigen-faces. The projection of images in this compressed subspace allows for easy comparison of images with the images from the database [3].

- Acquire an initial set of N face images (training images).
- Calculate the Eigen-face from the training set keeping only the M images that correspond to the highest eigen values. These M images define the "face-space". As new faces are encountered, the "Eigen-faces" can be updated or recalculated accordingly.
- Calculate the corresponding distribution in M dimensional weight space for each known individual by projecting their face images onto the "face space".
- Calculate a set of weights projecting the input image to the M "Eigen-faces".
- Determine whether the image is a face or not by checking the closeness of the image to the "face space".
- If it is close enough, classify, the weight pattern as either a known person or as an unknown based on the Euclidean distance measured.
- If it is close enough then cite the recognition successful and provide relevant information about the recognized face from the database which contains information about the faces.

Mathematically, it can be explained as given below. Assume $(X_1, X_2, X_3, \ldots, X_m)$ is a set of M train set from N face images arranged as column vector. Average face of set can be defined as:

$$\varphi = \frac{1}{M}\sum_{n=1}^{M} X_n$$

Each face differs from the average by vector

$$\emptyset_i = X_i - \varphi$$

When applied to PCA, this large set of vectors seeks a set of M orthogonal vectors U_n, which describes the distribution of data.

The K^{th} vector U_k is chosen such that

$$\vartheta_k = \frac{1}{M}\sum_{n=1}^{M}[U_k^T * \emptyset_n]^2$$

is maximum, applied to

$${U_k}^T U_k = \delta_{lk} = f(x) = \begin{cases} 1, if\ l = k \\ 0, otherwise \end{cases}$$

The vector U_k and scalar ϑ_k are the eigenvectors and eigenvalues respectively of the covariance matrix

$$C = \frac{1}{M}\sum_{n=1}^{M} \emptyset_n \emptyset_n^T$$
$$= AA^T$$

Where

$$A = [\emptyset_1, \emptyset_2, \ldots, \emptyset_M]$$

B. Linear Discriminant Analysis (LDA)

LDA also known as Fisher's Discriminant Analysis, is another dimensionality reduction technique. It is an example of a class specific method i.e. LDA maximizes the between – class scattering matrix measure while minimizes the within – class scatter matrix measure, which make it more reliable for classification. The ratio of the between – class scatter and within – class scatter must be high[5].

Basic steps for LDA:

Calculate within - class scatter matrix S_W:

$$S_W = \sum_{j=1}^{C}\sum_{i=1}^{n_j}\left(X_i^j - \mu_j\right)\left(x_i^j - \mu_i\right)^T$$

Where X_i^j is the i^{th} sample of class j is μ_j is the mean of class j, C is the number of classes, N_j is the number of samples in class j. Calculate between-class scatter matrix S_B:

$$S_B = \sum_{j=1}^{C}\left(\mu_j - \mu\right)\left(\mu_j - \mu\right)^T$$

Where μ represents the mean of the classes. Calculate the eigenvectors of the projection matrix

$$W = Eig(S_W^{-1} S_B)$$

Each and every test image is projected to the same subspaces and compared by the training images.

C. Independent Component Analysis (ICA)

Generalization View of the PCA is known as ICA. It minimizes the second order and higher order dependencies in the input and determines a set of statistically independent variables or basis vectors. Here we are using architecture I which finds statistically independent basis images[6]. Basic steps for ICA[7]:

Collect X_i of n dimensional data set X, i = 1, 2, 3… M. Mean corrects all the points: calculate mean and subtract it from each data point, − Calculate the covariance matrix:

$$C = (X_i - M_x)(X_i - M_x)^T$$

The ICA of X factorizes the covariance matrix into the following form:

$$C = F\Delta F^T$$

Where

Δisa diagonal real positive matrix.

F transforms the original data X into Z such that the components of the new data Z are independent: X=FZ [3].

This section describes the algorithms of statistical analysis for image data. The next section includes the detailed discussion of advantages and limitations of PCA, LDA and ICA algorithm.

IV. DISCUSSION

In this section we discuss, where these methods are beneficial and also the limitation of these methods, to be adopt in implementation.

Principal component analysis (PCA)

Principal component analysis (PCA) is a standard tool in modern data analysis - in diverse fields from neuroscience to computer graphics - because it is a simple, non-parametric method for extracting relevant information from confusing data sets. With minimal effort PCA provides a roadmap for how to reduce a complex data set to a lower dimension to reveal the sometimes hidden, simplified structures that often underlie it [8].

Importance of PCA is manifested by its use in so many different fields of science and life. PCA is very much used in neuro-science, for example. Another field of use is pattern recognition and image compression, therefore PCA is suited for use in facial recognition software for example, as well as for recognition and storing of other biometric data. Many IT related fields also use PCA, even artificial intelligence research. According to Jolliffe (2002) PCA is also used in research of agriculture, biology, chemistry, climatology, demography, ecology, food research (?), genetics, geology, meteorology, oceanography, psychology, quality control, etc. But in this PCA is a special case of Factor Analysis that is highly useful in the analysis of many time series and the search for patterns of movement common to several series (true factor analysis makes different assumptions about the underlying structure and solves eigenvectors of a slightly different matrix). This approach is superior to many of the bivariate statistical techniques used earlier, in that it explores the interrelationships among a set of variables caused by common "factors," mostly economic in nature [1].

PCA is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. A primary benefit of PCA arises from quantifying the importance of each dimension for describing the variability of a data set. PCA can also be used to compress the data, by reducing the number of dimensions, without much loss of information.

When using principal component analysis to analyze a data set, it is usually possible to explain a large percentage of the total variance with only a few components. Principal components are selected so that each successive one explains a maximum of the remaining variance, the first component is selected to explain the maximum proportion of the total variance, the second to explain the maximum of the remaining variance, etc. Therefore, the principal component solution is a particularly appropriate test for the existence of a strong market factor.

PCA is completely nonparametric: any data set can be plugged in and an answer comes out, requiring no parameters to tweak and no regard for how the data was recorded. From one perspective, the fact that PCA is non-parametric (or plug-and-play) can be considered a positive feature because the answer is unique and independent of the user.

Limitations in PCA occur mainly due to the previously mentioned main assumptions and the data at hand. PCA is not a statistical method from the viewpoint that there is no probability distribution specified for the observations. Therefore it is important to keep in mind that PCA best serves to represent data in simpler, reduced form. It is often difficult, if not impossible, to discover the true economic interpretation of PCs since the new variables are linear combinations of the original variables. In addition, for PCA to work exactly, one should use standardized data so that the mean is zero and the unbiased estimate of variance is unity:

$$Z_i = \frac{x_i - \mu_x}{\sigma_x}$$

Where$Z_i = i^{th}$standardized variable.

This is because it is often the case that the scales of the original variables are not comparable and that (those) variable (variables) with high absolute variance will dominate the first principal component.

There is one major drawback to standardization, however. Standardizing means that PCA results will come out with respect to standardized variables. This makes the interpretation and further applications of PCA results even more difficult. The mission when using PCA is often to get rid of correlation and interdependence of variables.

PCA succeeds in getting rid of second order dependences, but it has trouble with higher-order dependencies. This problem might be solved by using kernel PCA or independent component analysis. The fact that PCA is agnostic to the source of the data is also a weakness.

Linear Discriminant Analysis

Eigen value = sum of squared correlations of the discriminant function scores with the p original variables canonical correlation varies from 0 to 1 never taking on negative values More easily interpreted than an eigen value, though, is a direct expression of the proportion of between-group separation that is provided by each discriminant function. This proportion is computed by dividing the eigen value for a given discriminant function by the sum of the eigen values for all of the discriminant functions [9].

In SPSS the loadings are called the Canonical Structure Matrix. In SPSS the raw canonical coefficients are used to form a discriminant function that can be used to compute discriminant function scores for each person. The means of these scores form the centroids in the plots.

Advantages of Discriminant Analysis

1. Multiple dependent variables
2. Reduced error rates
3. Easier interpretation of Between-group Differences: each discriminant function measures something unique and different.

Disadvantages of Discriminant Analysis

Interpretation of the discriminant functions: mystical like identifying factors in a factor analysis Assumptions:

1. Each discriminant function formed is distributed normally in each group being compared.
2. Each discriminant function is assumed to show approximately equal variances in each group.
3. patterns of correlations between variables are assumed to be equivalent from one group to the next
4. the relationships between variables are assumed to be linear in all groups
5. no dependent variable may be perfectly correlated to a linear combination of other variables (Multi-co-linearity)
6. Discriminant analysis is extremely sensitive to outliers.

Interpretation of discriminant functions:

1. Begins with a series of univariate tests to determine which of the original dependent variables have contributed to the overall significance of the discriminant functions.
2. A discriminant function can be interpreted by determining which groups it best separates.
3. Correlations between a discriminant function and the original dependent variables can reveal what conceptual variable the discriminant function represents.

Independent Component Analysis (ICA)

This paper surveyed contrast functions and algorithms for ICA. ICA is a general concept with a wide range of applications in neural computing, signal processing, and statistics. ICA gives a representation, or transformation, of multidimensional data that seems to be well suited for subsequent information processing. This is because the components in the representation are 'as independent as possible' from each other, and at the same time 'as non-Gaussian as possible'. The transformation may also be interesting in its own right, as in blind source separation [10].

In the discussion of the many methods proposed for ICA, it was shown that the basic choice of the ICA method seems to reduce to two questions. First, the choice between estimating all the independent components at the same time, and estimating only a subset of them, possibly one-by-one. Most ICA research has concentrated on the first option, but in practice, it seems that the second option is very often more interesting, due to computational and other considerations. Second, one has the choice between adaptive algorithms and batch-mode (or block) algorithms. Again, most research has concentrated on the former option, although in many applications, the latter option seems to be preferable, again for computational reasons.

In spite of the large amount of research conducted on this basic problem by several authors, this area of research is by no means exhausted. It may be that the estimation methods of the very basic ICA model are so developed that the probability of new breakthroughs may not be very large.

V. CONCLUSION

The given paper is an evaluation of the feature extraction and selection technique. Where in order to find most optimum method is ICA, LDA and PCA statistical method is analysed. Due to literature collection and studied research article that is found that LDA is most optimum technique for the object recognition purpose, because of the proposed model includes more than one class definitions for a single object.

In near future a complete object recognition model is proposed and implemented using java technology and results are simulated.

REFERENCES

[1] Tuan Anh, Sander Mägi, "Principal Component Analysis", FinalPaperinFinancial Pricing, NATIONALCHENGKUNGUNIVERSITY, 6/17/2009

[2] S. Balakrishnama, A. Ganapathiraju, "LINEAR DISCRIMINANT ANALYSIS - A BRIEF TUTORIAL", Department of Electrical and Computer Engineering Mississippi State University, Box 9571, 216 Simrall, Hardy Rd. Mississippi State, Mississippi -39762

[3] Ajeet Singh, BK Singh and Manish Verma, "Comparison of HGPP, PCA, LDA, ICA and SVM", VSRD-IJEECE, Vol. 2 (4), 2012, 179-188

[4] Vapnik N. (1995): The Nature of Statistical Learning Theory, Springer.

[5] Zhang Baochang and et al (2007): Histogram of Gabor Phase Patterns (HGPP). A Novel Object Representation Approach for Face Recognition, IEEE Transactions on Image Processing, vol. 16, No.1, pp 57-68.

[6] Bartlett M. S., Movellan J. R., and Sejnowski T. J. (2002): Face Recognition by Independent Component Analysis," IEEE Transactions on Neural Networks, vol. 13, pp. 1450-1464.

[7] Martinez A.M., Kak A.C. (2001): PCA versus LDA, IEEE Trans. Patt. Anal. Mach. Intell. 23 (2) 228–233.

[8] Jonathon Shlens, "A Tutorial on Principal Component Analysis", Center for Neural Science, New York University New York City, NY 10003-6603 and Systems Neurobiology Laboratory, Salk Insitute for Biological Studies La Jolla, CA 92037 (Dated: April 22, 2009; Version 3.01)

[9] Notes on Discriminant Analysis, http://people.eku.edu/falkenbergs/psy862/notes%20on%20discriminant%20analysis.html

[10] AapoHyvärinen, "Survey on Independent Component Analysis", Neural Computing Surveys, Vol. 2, pp. 94-128, 1999.

[11] G.Karthikeyan And Radharamansahoo, "INDEPENDENT COMPONENT ANALYSIS BASED BLIND SIGNAL SEPARATION FOR MIXED SPEECH SIGNAL" ,Department of Electrical Engineering National Institute of Technology, Rourkela