PCA BASED CLASSIFICATION OF SINGLE-LAYERED CLOUD TYPES

Imran Sarwar Bajwa
&
S. Irfan Hyder
PAF-Karachi Institute of Economics & Technology.
Imransbjawa@yahoo.com : hyder@pafkiet.edu.pk

Abstract

The paper presents an automatic classification system, which discriminates the different types of single-layered clouds using Principal Component Analysis (PCA) with enhanced accuracy as compared to other techniques. PCA is an image classification technique, which is typically used for face recognition. PCA can be used to identify the image features called principal components. A principal component is a peculiar feature of an image. The approach described in this paper uses this PCA capability for enhancing the accuracy of cloud image analysis. To demonstrate this enhancement, a software classifier system has been developed that incorporates PCA capability for better discrimination of cloud images. The system is first trained by cloud images. In training phase, system reads major principal features of the different cloud images to produce an image space. In testing phase, a new cloud image can be classified by comparing it with the specified image space using the PCA algorithm.

Keywords: PCA, single cloud type recognition, principal components, eigenvectors.

1. Introduction

Weather forecasting applications use various pattern recognition techniques to analyze clouds' information and other meteorological parameters. Neural Networks is an often-used approach [3], [13] for image processing. Some statistical methodologies like FDA [4], RBFNN [1] and SVM [12] are also being used for image analysis. These methodologies require more training time and have limited accuracy of about 70% [11]. This level of accuracy often degrades classification of clouds, and hence the accuracy of rain and other weather predictions is reduced [15].

Principal Component Analysis (PCA) treats each image as an entity. A set of images is needed, which defines a *class* based on the core feature, derived from those images [6]. A single *class* can cover a certain number of images. The number of images in a *class* can be ineffective but the image quality can invariantly impact the overall image analysis results and consequences. PCA is used to avoid computational intense calculations. Its use results in fast and relatively more accurate inferences [7]. PCA is a way of identifying patterns in data and expressing the data in a way that highlights its similarities and differences. PCA strives to identify relatively fewer "features" or components that as a whole represent the full object state and hence are appropriately termed "Principal Components". Thus, principal components extracted by PCA implicitly represent all the features. However, these abstracted features may or may not include a specific feature [5].

Better accuracy in cloud classification means accurate categorization of clouds according to high, mid and low levels. These high, mid and low-level clouds are further classified in their particular sub classes illustrated in Section 3.3. PCA can easily handle a large amount of data due to its capability of reducing data dimensionality and complexity, thus getting better results [4]. The algorithm provides a more accurate cloud classification that yield better and concise forecasting of rain.

1.1. PCA and Eigenvectors

Training procedure in any classification system is significant and can be beneficial. Using various algorithms training can be performed. Algorithm used in current image classifier system for training is principal component analysis, whose major emphasis is to locate and depict the principal features of the given sample image [12].

The Eigenvectors is a key capability used in PCA analysis algorithm. Eigenvectors are defined to be a related set of spatial characteristics [6] that computer uses to recognize a specific cloud type. PCA technique uses training and testing sets of images. Eigenvectors of the covariance matrix is computed from the training set of images. These eigenvectors represent the principal components of the training images [7]. These eigenvectors are often ortho-normal to each other. In the context of clouds classification, these eigenvectors would form the cloud space. They may not correspond directly to any cloud feature like height, width and density. When the eigenvectors are displayed, they look like a ghostly cloud. They can be thought of as a set of features that together characterize the variation between cloud images.

Cloud Detection consists in locating a cloud in complex scenery, by locating and cutting it out. Some methods search elliptical and polygonal forms [2], others seek the texture and color of the clouds and still others seek the patterns and boundaries of the cloud [3].

PCA is used abundantly for image analysis and classification purposes [6] because it is a simple, non-parametric method of extracting relevant information from confusing data sets [1]. PCA extracts relevant features from data by performing an orthogonal transformation. PCA also provides assistance to reduce a complex data set to a lower dimension [7]. This paper demonstrates that PCA is relatively better than other techniques in discriminating different types of single-layered cloud images.

2. Methodology

2.1 System's general Architecture

The developed classifier system discriminates the single-layered cloud types. It carries out classification in five modules: image acquisition, detection, extraction of related attributes, comparison of these attributes and finally classification as shown in Fig. 1.

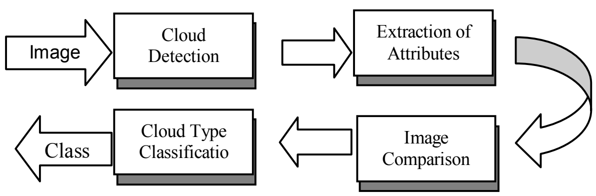

Fig 1: General architecture of the cloud types classification system

2.1.1. Image acquirer

This module helps to acquire new image. The image can be acquired through different sources e.g. digital camera. Images for testing and training phases are converted to 256-bit Gray color image. Images are also scaled to 50 x 50 ratio. This ratio can vary from 30 x 30 to 50 x 50.

2.1.2. Cloud Detection

This module detects the presence of the cloud fragments in the images. In this module it is specified that rather the images contain the clouds or not.

2.1.3. Extraction of Attributes

This module identifies the various patterns in data. Cloud attributes are extracted from images using Principal Component Analysis algorithm.

2.1.4. Image Comparison

This module compares the principal features of the test image with the image-space of already given images in training set. After matching it infers that rather image is recognized or not.

2.1.5. Cloud Type Classification

This module finally detects and classifies the cloud type. Images are classified to their respective types on the basis of the matching inferences provided by previous module.

2.2. PCA based Extraction of the Attributes

The system presented in this work exemplifies the concept of Eigenvectors. These eigenvectors are a small group of characteristics extracted by the designed classifier system using PCA. PCA is a two-phase algorithm consisting of Training followed by Recognition.

2.2.1. Training

Training phase constructs an image-space, called a cloud space, which is later required for classification in testing phase. In training phase, the classifier system is trained by using sample data input. If it is required, output pattern can be enhanced and improvised by retraining the system by more refined and conspicuous data. Training is performed using *n* images in the following 6 steps:

STEP 1:

Each sample image is converted into a row vector. A row vector can be constructed by concatenating each row with first row in sequence. As in fig-2 a *m* x *n* matrix is converted into a single row *1* x *mn* vector *X*.

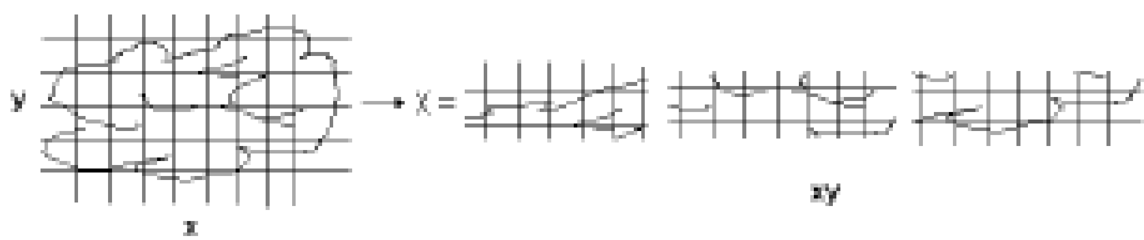

Fig 2: A row vector representation of a 2-D cloud image

STEP 2

The row vector matrix is constructed by combining together the row vectors of *n* cloud images. X_i is a row vector of a sample image *i,* where *i* = 1 .. n.

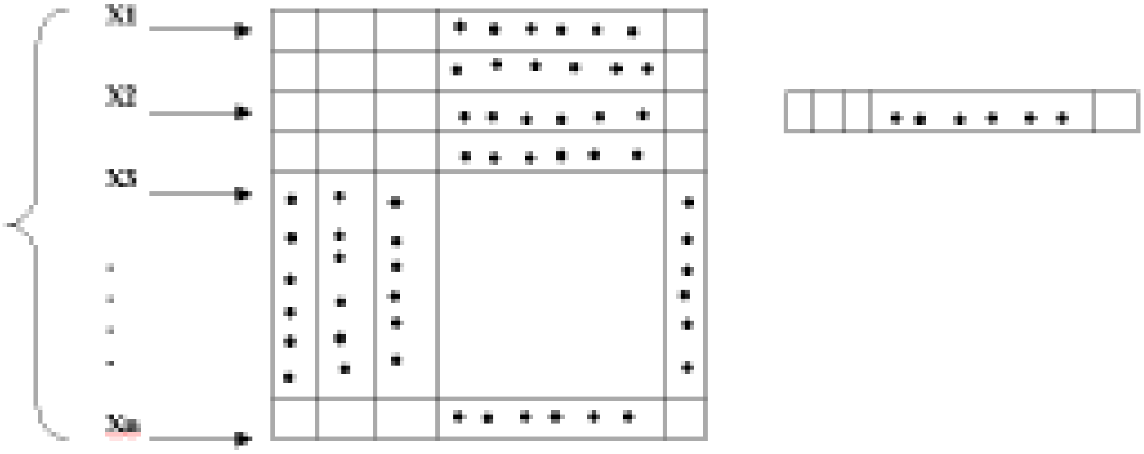

Fig 3: Whole cloud distribution. Row vector of X_i

STEP 3

A mean cloud vector Ψ of n row vectors is calculated to extract required principal features.

$$\Psi = (1/n)\ \Sigma X_i \qquad \textit{where } i = 1 .. n$$

STEP 4

A new matrix Φ is constructed by subtracting mean cloud vector Ψ from each cloud image X of the training set.

$$\Phi_i = X_i - \Psi$$

STEP 5 *

A data covariance matrix C is calculated by multiplying matrix Φ with its transpose matrix Φ_t.

$$C = \Phi_t\, \Phi$$

STEP 6 *

20 highest valued eigenvectors are then picked to make an image space from the resultant covariance matrix C.

* (*Step5,6 are performed using Matlab 5.6*)

2.2.2. Recognition

In testing phase, each new image is analyzed and its principal features are located. Then these principal features are compared with the principal features of image-space. If some match is found there, then the image is classified according to the previously defined rules. Recognition or testing phase is performed in the following two steps.

STEP 1

A new cloud image is categorized by calculating projection Ω on image-space by

$\Omega = U_i * (Z - \Psi)$

Where U_i is image-space and Z is the new Image

STEP 2

If threshold Φ matches with one of the thresholds in image space then cloud recognition occurs and the particular cloud type is specified.

$\Phi i \quad = 1/k \max (\Omega i \ -\Omega j) \quad \text{where } (i,j=1 \ldots.. n)$

3. Experiments

A series of experiments were done using the developed classified system to evaluate its accuracy. Experiments were performed using following steps:

1- Data Collection
2- Normalizing Cloud Images
3- Define Classes
4- Cloud Type Classification
5- Evaluate Accuracy
6- Comparison with other Technologies

3.1- Data Collection

Image data of cloud's different types was obtained for training purposes. This data is available from different sources. Ground-based cloud images have been used in this experiment. These images of the general and sub-cloud types are available at different websites of world's major weather forecasting organizations [13], [14] , [15]. Overlapping sets of training images and testing images were used for the experiment. Global daytime cloud images are used in development and implementation aspects of a principal component analysis classification system. The designed image classifier system is used to find the presence of clouds and classification of single-layer clouds in cloud images.

3.2- Normalizing Cloud Images

Images for testing and training phase are of 256-bit Gray color image and are overlapping. If the acquired image is not in specified bitmap format then it is converted into required format. The system obtains the image in the form of BMP of JPEG format.

Acquired Image was of size 50 x 50 pixels for processing in the designed system. But this ratio can be tuned from 30 x 30 to 50 x 50. This module gets the image in integer or short co-ordinate i.e. perform scaling at 50 x 50 scale.

3.3 Define Classes

An efficient and effective image classifier system often consists of a defined set of classes. These precisely defined classes are well separated by a set of features that are typically derived from the multi-dimensional radiometric image data. The selection of classes is often influenced by desired application and classes may be complicated. In this research, there are four defined general classes and these general classes are further divided into sub classes and they are

1- Clear sky
2- Low-level clouds
  i)- cumulus
  ii)- Stratocumulus
  iii)- Stratus
3- Mid-level clouds
  i)- Altocumulus
  ii)- Nimbostratus
  iii)- Altostratus
4- High-level clouds
  i)- Cirrus
  ii)- Cirrostratus
  iii)- Cirrocumulus

3.4- Cloud Type Classification

Two types of satellite images have been used, first as training image and other images with clouds for testing. Comparing the individual pixel values within 50 x 50 array with a clear sky images depicts cloud fragments in a sample image. Often the array of 32 x 32 array is used in conventional image recognition applications. As the greater number of pixels can immensely affect the memory usage so array of smaller range is preferred. But this procedure also affects the overall image processing accuracy. But PCA handles images so conveniently that an array of greater range may be used to get still higher accuracy.

If the cloud matches with the existing collected data then the program will display as match is found. It displays cloud's general type as low, mid or high. Program has also the capacity of prescribing the cloud's sub type and also describing the properties of each cloud sub type.

3.5- Evaluate Accuracy

To test the accuracy of the designed system images with clear sky and images with all types of single-layered clouds are used. 17 Clear sky images were used for testing and all images were successfully categorized. 36 Images of single-layered cloud types were used and showed results with high accuracy. A matrix of results of testing images is shown below.

Table 1. Testing results of different cloud type images

A matrix representing classification accuracy test (%) for cloud free and single-layered cloud types is constructed. Classification inferences of Principal Component Analysis for different cloud types are shown in the matrix. Overall classification accuracy for single-layered clouds is determined by dividing number of correctly classified samples by the total number of samples. An accuracy test (%) table is shown here.

Classes	**Clear sky**	**Low-level**	**Mid-level**	**High-level**
Clear sky	17	0	0	0
Low-level	0	34	1	1
Mid-level	0	1	32	3
High-level	1	1	3	31

Table 2. Total Accuracy = 92.3%

3.6- Comparison with other Techniques

Various classification techniques and algorithms are used for image classification. Each technique has its own respective accuracy level. The derived results using principal component analysis are compared with the results of other technologies used for cloud classification. Different technologies provide with different level accuracies.

Classes	**Clear sky**	**Low-level**	**Mid-level**	**High-level**
Clear Sky	100	---	---	0.5
Low-level	---	94.1	0.5	0.2
Mid-level	---	0.2	88.9	0.8
High-level	0.3	0.3	0.8	86.2

Results show that PCA, relative to other statistical techniques, is more accurate [table. 3]. Other statistical techniques include Fuzzy Logic based systems that give 84% accuracy [5] but Fuzzy systems are dependent on the appropriateness of the initial categories defined i.e. much effort is needed for domain knowledge and efficiency issues. Neural networks demand intense domain knowledge and intuition for representation otherwise suffer from divergent training sessions and inaccurate results [11].

PCA is the image classification technique, which provides higher accuracy up to 90%. Statistics show that PCA based image classifier system is a better classifier than other used techniques. A comparison of PCA with other techniques is also given below.

Table 3. Accuracy comparison in different techniques

4. Conclusion & Future Work

PCA is an efficient identifier in terms of time and provides better accuracy in cloud image recognition. A PCA-based system provides high speed processing with relatively better accuracy. PCA also easily handles a large amount of data due to its capability of reducing data dimensionality and complexity. PCA algorithm provides a more accurate cloud classification that infers better and concise forecasting of rain. Probably, the more long-term weather forecasting is also possible.

In this report only one type of cloud has been addressed and that is single-layered clouds. Other type is Multi-layered clouds. Multi-layered clouds are 60 % to 65% of total clouds. So their identification and classification is also a significant task.

Future goal of the research is to analyze all cloud types using satellite image. Satellite image contains clouds of all types and there is very complex information. For simple cloud recognition, co-variance matrix has been used. To cover the whole scenario (all cloud types) another type of matrix named, co-exist matrix can be used. There is also need of generalizing the algorithm. This process may need much more effort. Furthermore, the accuracy estimations can be further improved using a more carefully selected disjoint sets of image data for the training set and the testing set.

Technology Name	Accuracy	Error Ratio
PCA (Principal Component Analysis)	92.30%	0.9%
NMF (Non-Negative Matrix Factoriz.) [8]	69.94%	8.1%
BPNN (Back Propagation NN) [9]	71.80%	
RBFNN (Radial Basis Function NN) [1]	73.20%	7.3%
SVM (Super Vector Machine) [12]	84.11%	
Fisher Discriminant Analysis [4]	64.00%	
FLNN (Fuzzy Logic Neural Networks) [5]	81.00%	3.4%
Wavelet Transforms [6]	78.30%	3.9%
Probabilistic Neural Networks [2]	86.01%	
K-SOM (K-Self Organizing Maps) NN [10]	80.00%	

References

[1]- Su Hongtao, David Dagan Feng, Zhao Rong-chun , 1997. Face Recognition Using Multi-feature .... and Radial Basis Function Network.

[2]- Bankert, R. L., 1994: Cloud classification of AVHRR imagery in maritime regions using a probabilistic .neural network. *Journal of Applied Meteorology.*, 33, 909–918. Boulder, CO 80307.

[3]- Barsi, A., Heipke, C., Willrich, F., 2002, Detecting road junctions by Artificial Neural Networks – JEANS, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science* (34) 3B, pp. 18-21

[4]- Seong-Wook Joo, December 2003. Face Recognition using PCA and FDA with intensity normalization.

[5]- Bryan A. Baum, Vasanth Tovinkere & Jay Titlow, Ronald M. Welch, 1997. Automated Cloud Classification of Global AVHRR Data Using a Fuzzy Logic Approach. *Journal of Applied Meteorology. pp* 1519-1539.

[6]- Chi-Fa Chen, Yu-Shan Tseng and Chia-Yen Chen, 2003. Combination of PCA and Wavelet Transforms for Face Recognition on 2.5D Images. *Conf. of Image and Vision Computing* '03 26-28 November 2003

[7]- Cristina Conde ,Antonio Ruiz and Enrique Cabello, 2003. PCA vs Low Resolution Images in Face Verification. *Proceedings of the 12th International Conference on Image Analysis and Processing* (ICIAP'03).

[8]- David Guillamet, Bernt Schiele, and Jordi Vitri. Analyzing Non-Negative Matrix Factorization for Image Classification. *Conference on Pattern Recognition ICPR 2002*, Quebec, Canada, August 2002

[9]- Kishor Saitwal, r. Azimi Sadjadi and Donald Rinki, 1997. A multi-channel temporarily adaptive system for continuous cloud classification from Satellite Imagery.

[10]- Bin Tian, Mamood R. AzimiSadjadi, Thomas H. Vonder and Donald Rienki, 1999. Neural Network based cloud classification on satellite Imagery using textural features

[11]- Simon Haykin , *McMaster University, Ontario Canada*1998, Neural Networks a Comprehensive Foundation , 2/E, *Prentice Hall publishers*

[12]- D. Cao, O. Masoud, D. Boley, N. Papanikolopoulos, 2004. Online Motion Classification using Support Vector Machine, IEEE 2004 *International Conference on Robotics and Automation*, April 26 - May 1.

[13]- VanderZwaag, B.J., Slump, C.H. and Spaanenburg, L. (2002) Analysis of neural networks for edge detection, Proceedings. *Pro-Risc*'02 (Veldhoven) pp. 580-586

[14]- http://www.noaa.gov/

[15]- http://www.eumetsat.de/en/area2/cgms

[16]- http://www.nottingham.ac.uk/meteosat/