www.icgst.com

Feature Based Image Classification by using Principal Component Analysis

Imran S. Bajwa[1], M. Shahid Naweed [1], M. Nadim Asif [1], S. Irfan Hyder[2]

1 Department of Computer Science & IT, The Islamia University of Bahawalpur, Pakistan
2 PAF-KIET, Karachi, Pakistan.

imransbjawa@yahoo.com, shahid_naweed@hotmail.com, nasif@softresearch.org, hyder@pafkiet.edu.pk

Abstract

Classification of different types of cloud images is the primary issue used to forecast precipitation and other weather constituents. A PCA based classification system has been presented in this paper to classify the different types of single-layered and multi-layered clouds. Principal Component Analysis (PCA) provides enhanced accuracy in features based image identification and classification as compared to other techniques. PCA is a feature based classification technique that is characteristically used for image recognition. PCA is based on principal features of an image and these features discreetly represent an image. The used approach in this research uses the principal features of an image to identify different cloud image types with better accuracy. A classifier system has also been designed to exhibit this enhancement. The designed system reads features of gray-level images to create an image space. This image space is used for classification of images. In testing phase, a new cloud image is classified by comparing it with the specified image space using the PCA algorithm.

Keywords: *Feature identification, Multi-layered cloud types recognition, principal components, eigenvectors, weather prediction*

1. Introduction

Satellites are major source of data for forecasting weather applications. Weather forecasting applications require better processing speed and also accurate analysis of data (images). Here analysis is comprised of a number of steps: image pre-processing, image enhancement, image identification, image classification, etc. Various image processing and pattern matching and recognition techniques and methodologies are used to analyze image information. The used technologies can be differentiated into two halves: statistical and non-statistical approaches. Some statistical methodologies like FDA [4], RBFNN [1] and SVM [12] are incorporated for image analysis. In non-statistical techniques, Neural Networks is an often-used approach [3, 13] for image processing. Existing used methodologies have some issues: require more training time, exploit more processing 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]. We presented PCA approach for identification and analysis of single layered cloud-types [19]. PCA presented comparatively better results then previously used techniques. In this research, another variation of PCA, KPCA is also used for image identification.

1.1 PCA vs KPCA

Principal Component Analysis (PCA) [1] extracts principal features of an image. These features are integrated in a single *module* or *class* [6]. These features can principally differentiate among various input images. This technique produces results in fast and relatively more accurate manner [7].

On the other hand, KPCA [18] is Kernel Principal Component Analysis (Kernel PCA) is another variation of PCA. KPCA is a method of non-linear feature extraction, closely related to methods applied in Support Vector Machines (SVMs) [19].

PCA has ability 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.

1.2 Principal Features

Principal features in PCA are represented by *Eigenvectors*. The Eigenvectors are defined to be a related set of spatial characteristics [6] of an image that a computer uses to identify and recognize a specific cloud type. 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. 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 [3] and still others seek the patterns and boundaries of the cloud. When the eigenvectors of these features 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.

In the following section of this paper, related work review and used methodology has been presented. The used algorithm has also been elaborated later in the section. Next section presents the experiments and the results with analysis. Last section of the paper describes the conclusion of the research work.

2 Literature Review

Cloud classification is the major research area in meteorology. Many researchers have done work in this area to classify different types of clouds. Multi-layered clouds tend to cover large areas and are indicated on a ground-based or satellite picture by an area of uniform brightness. On the other hand, single-layered clouds are usually formed by air being heated from below.

S. C. Ou and Y. Takano [13] designed a system for remote sensing of cirrus cloud parameters using advanced very-high resolution radiometer. Bankert, R. L. [2] classified clouds into one of ten output classes using a probabilistic neural network (PNN) applying on advanced very high resolution radiometer data of 16 pixel × 16 pixel sample areas. Bryan A. Baum [5] presented a fuzzy logic classification (FLC) methodology is to discriminate between clear sky and clouds in a 32 × 32 pixel array and to discriminate between single-layered and multilayered clouds within the sample. Cloud-top height was determined using ATSR data by Turner, P.J. [17]. The work estimated more accurate cloud top height from satellite images for weather forecasting and analyzing climate changes and they used Statistical Methods. This article presents a facial expressions classification experiment using neural networks.

Su Hongtao presented a classification system [1] that was based on attributes extraction from human face images using the principal component analysis (PCA) technique. Well-framed images were used in order to simplify the face detection on the image. Two different models of neural networks were applied as classifiers: back-propagation and RBF networks. For classification of single-layered cloud types another system [19] was presented that was based on PCA. The system had better accuracy in classification results. Now the same project is extended to classify multi-layered clouds as well.

3. Designed System's Methodology

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. Following are some major phases of the designed system.

a. Image Acquisition - 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.

b. Detection of Cloud Fragments - 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.

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

d. 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.

e. 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. First image is classified as single-layered or multi-layered cloud image and then afterwards the sub type is identified. Figure 2.1 represents the described architecture of the designed system.

4. Feature Extraction using PCA

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 [19].

- Training Phase
- Recognition Phase

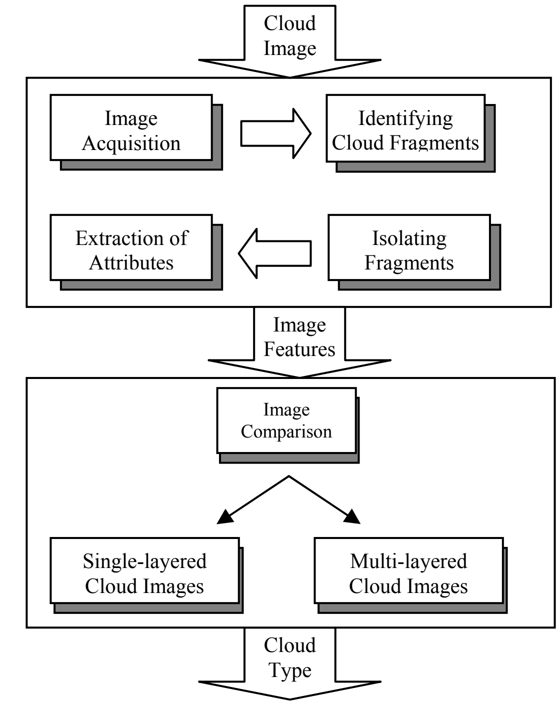


Figure 2.1: Architecture of the cloud types classification system

4.1. Training Phase

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.

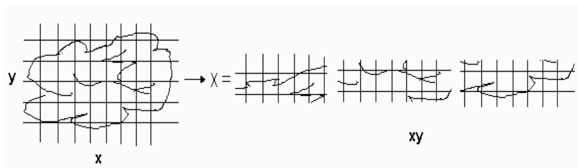


Figure 2.2: A row vector representation of a 2-Dimensioanl cloud image

A training image of 50 x 50 pixels is taken and converted into a 1-row vector of length 2500 x 1. The procedure has been explained in figure 2.0.

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.

1-row vectors of 10 images are combined to make a 2-D array of 10 x 2500

STEP 3

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

$$\Psi = (1/n) \sum Xi \quad where\ i = 1\ .....\ n$$

A mean vector (1 x 2500) is taken of 2-D array of 10 x 2500.

STEP 4

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

$$\Phi i = Xi - \psi$$

Mean vector (1 x 2500) is subtracted from matrix of step 2 to make a new 2-D array of 10 x 2500

STEP 5*

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

$$C = \Phi t \text{ x } \Phi$$

Covariance matrix of step 4 is calculated using Matlab 6.5

STEP 6*

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

In this experiment, 20 highest values of each image are taken to result a final matrix that is used for recognition

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

4.2. Recognition Phase

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 on image-space by

$$\Omega = U_i \text{ x } (Z - \psi)$$

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

STEP 2

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

$$\Phi i = 1/K \max (\Omega i - \Omega j) \quad where\ (i, j = 1, \ldots\ldots n)$$

If threshold of any recognition image matches with the training work space, recognition occurs

STEP 3

A set of rules are defined to classify the corresponding type of the matched image. Each training image has attached related information in coded form e.g. main type of cloud, sub-type of cloud, etc. This information is coded at the time of training.
The cloud image is classified into specific type according to the specified rules.

5. Experiments and Analysis

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

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.

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 two major classes.

1. Single-layered Clouds
2. Multi-layered Clouds

Single-layered cloud images are further classified into following sub-categories.

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

Normalizing Cloud Images

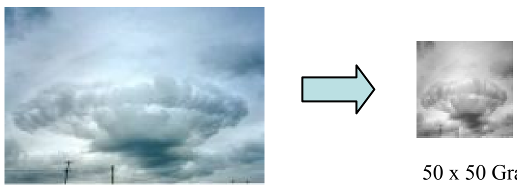
Images for testing and training phase are of 256-bit Gray color image and are overlapping. If the acquired image is not in specified bitmap format then it is converted into required format. The system obtains the image in the form of BMP of JPEG format.
Acquired Image was of size 50 x 50 pixels for processing 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.

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.

Details of Experiments

A software system was designed to practically perform the experiments and check the affectivity of the PCA algorithm. The code of designed system was written in Visual C language. First phase converts normal 24-bit image into 8-bit gray-scale image.



Cumulus – Normal Coloured Image

50 x 50 Gray scale (8-bit) bmp image

Figure 3.1: Conversion to 24-bit to 8-bit gray-scale image

Following are the some examples of the single-layered cloud images that have been used for the training or recognition purpose.

Altostratus
(*Mid Level Cloud Image*)

Nimbostratus
(*Mid Level Cloud Image*)

Cirrus
(*Higher Level Cloud Image*)

Cirrocumulus
(*High Level Cloud Image*)

Stratus
(*Low Level Cloud Image*)

Altocumulus
(*Mid Level Cloud Image*)

Figure 3.2: 8-bit Converted images for training

First window is *load training* that is used for image acquisition. First of all a set of images that would be use in training are loaded in system, so that their principal features can be studied. Training can take place with out this step but this procedure make the process of training faster and causes minimum problems.

Figure 3.3: The Start Training window of the designed system

Window shown in figure 3.3 performs the preliminary steps involved in training. At this step actual training of the system is performed by extracting different principal features of the image. Then it is informed to the system that to which cloud type this image belongs and then the further sub-cloud type is also specified manually. This phase need an adequate training. The sufficiency of training would be specified by the variance of data. Higher the data variance in the sample images, higher the training would be needed.

A data file was created on the basis of first four steps of the training phase. Step 5 and step 6 were implemented in MATLAB 6.5 software. MATLAB comprehensively handles the complexity of steps 5 and 6. After receiving the processed file from MATLAB, the designed software system completes the remaining training procedure. Figure 3.4 shows working of image recognition module that performs actual matching of the images. If the match occurs then the inferences are sent to the next module for further classification. Here the co-variance matrix is generated and finally the highest valued eigenvectors are selected.

Figure 3.4: The *Complete Training* window of designed system

Third and final window of the designed system presents actual classification of the cloud images. This testing is performed on the basis of training performed and features specified by previous phase. This class simple compares the principal features of new testing image with the existing training image's features. On the basis of the various existing features of the cloud images the classifier software system displays the main type and sub type of that cloud.

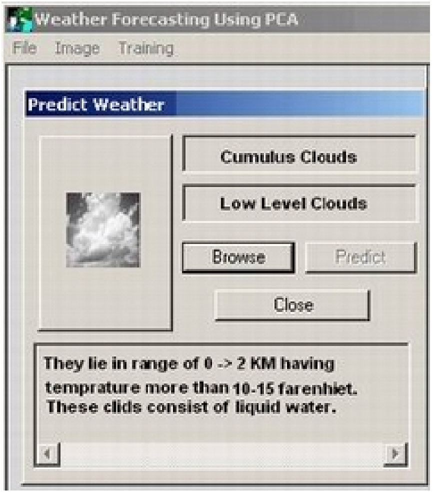


Figure 3.5: The *Testing* or *recognition* window of designed system

Evaluate Accuracy

To test the accuracy of the designed system images of different type were used. 20 Clear sky images were used for testing and all images were successfully categorized. 36 images of each single-layered and multi-layered cloud types were used and showed results with high accuracy. A matrix of results of testing images is shown below.

Classes	Clear sky	Single-layered	Multi-layered
Clear sky	20	0	0
Single-layered	0	33	3
Multi-layered	0	5	31

Table 3.1 - Testing results of different cloud type images

A matrix representing classification accuracy test (%) for cloud free and single-layered and multi-layered cloud types is constructed. 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	Single-layered	Multi-layered
Clear sky	100	0	0
Single-layered	0	91.6	2
Multi-layered	0	4	86.11

Table 3.2 - Average Accuracy = 88.85%

Comparison with other Techniques

Various classification techniques and algorithms are used for image classification. Every technique provide with 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.

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 itself 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].

Technology Name	**Accuracy Per.**	**Error Ratio**
PCA (Principal Component Analysis)	88.85%	1.2%
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 [7]	80.00%	

Table 3.3 - Accuracy comparison in different techniques

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. There is a comparison of PCA with other techniques is also given below.

6. Conclusion

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.

References

[1] Su Hongtao, David Dagan Feng, Zhao Rong-chun , 1997. Face Recognition Using Multi-feature and Radial Basis Function Network. *Visualisation 2002 of the ACS Conferences in Research and Practice in Information Technology (CRPIT),* Volume 22.

[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] Bin Tian, Mamood R. AzimiSadjadi, Thomas H. Vonder and Donald Rienki, 1999. Neural Network based cloud classification on satellite Imagery using textural features
[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] 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).
[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] S. C. Ou, K. N. Liou, W. M. Gooch, and Y. Takano, [1993] "Remote sensing of cirrus cloud parameters using advanced very-high resolution radiometer 3.7- and 10.9-m channels," *Appl. Opt.* Vol. 32, pp. 2171–2180.
[14] http://www.noaa.gov/
[15] http://www.eumetsat.de/en/area2/cgms
[16] http://www.nottingham.ac.uk/meteosat/
[17] Prata, A.J., Turner, P.J., [1997] "Cloud-top height determination using ATSR data" Remote Sensing Env., 59, 1, pp. 1-13.
[18] B. Schölkopf, A. Smola, K. R. Müller, [1998], "Nonlinear component analysis as a kernel eigenvalue problem", Neural Computation, 10:1299-1319, 1998
[19] Imran S. Bajwa, S. Irfan Hyder [2005], "PCA based Image Classification of Single-layered Cloud Types", Journal of Market Forces Vol.1 No.2, pp. 3-13
[20] B. Schölkopf, C. J. C. Burges, A. J. Smola, [1999] "Advances in Kernel Methods - Support Vector Learning", MIT Press, Cambridge, MA, 1999.

Author's Biography

Imran Sarwar Bajwa is Assistant Professor in the Islamia University of Bahawalpur.. He has been teaching and doing research in various universities in Pakistan and University of Coimbra, Portugal.

Prof. Dr. Irfan Hyder is president of PAF-KIET University in Pakistan. He has done his PhD from UT-Austin.