GIK Institute

Faculty of Engineering Sciences,

Assessment Type: Complex Engineering Problem / Course Project	**Course Name / Code:** Linear Algebra II / ES-304
Section: n/a	**Instructor:** Dr. Asad Mahmood
Semester: Fall 2023	**Weightage: 10%**

Concerned CLOs:

1. CLO4 - Analyze and solve applied engineering problem requiring tools from advanced linear algebra. PLO-4 (Investigation)
2. CLO5 - Efficiently work in a team to investigate and solve problems related to applied linear algebra. PLO-9 (Individual and Teamwork)

Title:

How to handle Large Datasets? → PCA comes to rescue!

1. Background:

Many satellites continuously roam around the earth for diverse applications. An important type of satellites is known as the **'Earth observation Satellites'** [1] which observe the earth from space using several sensors, such as cameras etc. This largely falls in the discipline of **'Remote Sensing'** [2]. The images from the cameras onboard the satellites can be used for several purposes, and one important application is known as **'land mapping'** [3], in which different areas on the land are mapped to different classes such as water, forests, crops, buildings, snow etc. Such information is useful in applications such as urban planning, water and crops monitoring, disaster management etc.

The earth observation satellites use different types of cameras to take pictures of the earth's land cover. One type of camera, which is commonly used in earth observation, is the multispectral camera which takes **multispectral images** [4] of different locations on earth. Multispectral images are similar in nature to the normal colored images which we use in our everyday life, however, instead of just 3 frequencies/spectral bands (Red, Green and Blue) which are in regular color images, there are a greater number of bands in a multispectral image (generally between 6 to 30 bands). The large number of bands allows one to distinguish between different land cover types (crops, forests, water, land, etc.) in a better manner as compared to using only the regular colored/RGB images. **Copernicus** is one of the world's largest programs for earth observation which is run by the European Union (EU). It makes use of **Sentinel satellites** to record images of the earth at various wavelengths from space. Landsat is another such program by US government. Some details about the Landsat program are also given in the introductory example of chapter 7 of your textbook. The Copernicus program from EU keeps an online repository of the

earth's images taken from Sentinel satellites. and allows anyone to download the images for any location using its online platform [7]. You will need to create an account on this platform to be able to download data from it. You can then use their *Earth Browser* to download data of any location on earth. Preferably use a location from Pakistan and use L2A (reflectance) data from Multispectral instrument on Sentinel. Details on how one can use the Earth Browser to download image/data are also provided on their documentation webpage [8].

2. Main Objective:

Your main objective in this Project/CEP assignment is to **perform principal component analysis (PCA)[1] [9] on a multispectral image and perform associated error analysis** using any appropriate platform/language (Matlab, Python etc.).

3. Tasks:

A. Develop understanding of multispectral/multidimensional images, download a multispectral, preferably L2A (reflectance), data/image of a location (could be any location of interest in Pakistan) from Sentinel. **Perform basic operations (visualize different bands, crop size in any dimension, concatenate bands etc.) with the image** in any suitable platform (Matlab, Python etc.). Note that images for any 2 groups should not be the same – this would be done by mentioning your chosen location on MS Teams). **(2 Marks)**

B. Develop understanding of the PCA technique, including how it helps in the processing (*dimensionality reduction*) of multidimensional images and how does PCA relates to the concepts from the course of Linear Algebra II. **Apply PCA to your image** to reduce the spectral dimension/'no. of spectral bands'. **(Note that you are NOT allowed to use any large function for PCA implementation e.g. pca(), svd() etc., but rather all individual steps involved in PCA application should be implemented using simpler functions, such as those used for calculation of eigenvalues/eigenvectors) (4 Marks)**

[1] A brief introduction about PCA is also given in chapter 7 of your textbook

C. **Perform an error analysis on the *dimensionality reduction* via the use of PCA**. Choose different numbers of Principal Components/reduced dimensions and calculate the corresponding error/'information loss' in each case, using an appropriate metric for information loss. **Based on your analysis, recommend a suitable number of Principal Components to be used**, such that there is an acceptable tradeoff between the data/dimension reduction and the associated information loss. **(2 Marks)**

D. **CEP report**, including individual contributions, which your team will use to give a viva on your CEP as well. **(1+1 = 2 Marks)**

4. Milestones and Deadlines:

Sr. No.	Milestone	Weightage	Due Date
1	**Milestone 1: CEP Group and Image Location selection**	n/a	**End of Week 1 after Mid-exams (November 17, 2023). Late submissions will be penalized.**
2	**Milestone 2: Source-code and CEP report submission**	2%	**Week 5 after Mid-exams (December 15, 2023). Late submissions will be penalized.**
3	**Milestone 3: CEP demo and viva**	8%	**Week 6 after Mid-exams (December 18-22, 2023)**

4.1. Milestone 1: CEP Group and Image Location selection

- Please **form a group of 4 students** and mention the names and reg. nos. of the students in the group as well as your chosen location of Sentinel Satellite image on **MS Teams by the due date for milestone 1.**
- The group will be assigned group no. which the instructor and the students will use in their CEP reports and viva.
- **Late submissions will be penalized**.

4.2. Milestone 2: Source-code and CEP report submission

- The **final submission must include the report as well as the source code. The report and source code as a single zip file will be submitted electronically via the MS Teams platform, and a hard copy of the report will be submitted by the due date to course TA by the due date**.
- **Late submissions will be penalized**. Similarly, failure to adhere to instructions in this handout will also result in a penalty.
- Source code must be properly formatted, should be easily executable and should be with comments for clarity regarding different lines of codes. Preferably use Jupiter notebooks, which show important intermediate results/images within the code as well.
- The report should be of standard format with no more than 5 pages in double-column format – **preferably use the IEEE Word template for conferences** [10].
- The report should include the sections of Title, Authors names, Reg. nos and group no., Abstract, Introduction, Individual sections on each CEP task and its methodology/implementation details and results, Task distribution (this section should be reasonably explained as there is a dedicated CLO regarding it in this assignment), Conclusion and References. Any item other than these can be put in the appendices.
- Please **follow good practices in report writing**, such as avoiding formatting and language errors, use of reasonable font type and size for headings and text, number and title figures and tables, avoid unnecessary/redundant or shallow discussions etc. You can go through some relevant references in this regard [11,12].

4.3. Milestone 3: CEP Demo and Viva

- Vivas will be arranged afterwards where each group will get a time slot to present their project and its understanding. **Each team-member will be asked questions related to the project and may be given a small task regarding any aspect of the project, and marks will be awarded individually**.
- Failure to register for a viva in a dates or **no-show for the viva will result in penalty**.

5. Tips (which will help you in your assignment):

- You may want to download a Sentinel image with a relatively smaller size.
- Initially, extracting a small portion from your downloaded image, and performing all the tasks on it may be more helpful.
- You may want to see how you can concatenate all the bands in one dataset so that you can reduce the spectral dimension using PCA.
- Reading up on PCA from the book and/or from the web in general will be helpful.
- Initially, performing the PCA on a generic (non-image) dataset, may be helpful to develop an understanding.

References:

1. https://en.wikipedia.org/wiki/Earth_observation_satellite
2. https://en.wikipedia.org/wiki/Remote_sensing
3. https://www.un-spider.org/links-and-resources/data-sources/daotm-landcover
4. https://en.wikipedia.org/wiki/Multispectral_imaging
5. https://en.wikipedia.org/wiki/Copernicus_Programme
6. https://www.copernicus.eu/en
7. https://dataspace.copernicus.eu/
8. https://documentation.dataspace.copernicus.eu/Applications/Browser.html
9. https://en.wikipedia.org/wiki/Principal_component_analysis
10. https://www.ieee.org/conferences/publishing/templates.html
11. https://www.sussex.ac.uk/ei/internal/forstudents/engineeringdesign/studyguides/techreportwriting
12. https://ias.ieee.org/wp-content/uploads/2023/06/2020-01-16_IET_Technical_Report_Writing_Guidelines.pdf

CEP Assignment for ES-304 (Rubric Based Marking) - Fall 2023

Sr. No.		V. Poor (1-2)	Poor (3-4)	Satisfactory (5-6)	Good (7-8)	Excellent (9-10)	Marks
1.	Obtain multidimensional images of interest and **perform basic operations** (bands visualization, cropping, concatenation etc.) on them. **(Marks=2)**	Unable to get required images and cannot perform basic operations on them	Able to get required images but cannot perform basic operations on them	Able to get required images and can perform only some of the basic operations on them	Able to get required images and can perform only most of the basic operations on them	Able to get required images and can perform all of the basic operations on them	
2.	Understanding of PCA, its relation to the course and **PCA application to the images using a suitable Software. (Marks=4)**	Cannot demonstrate an understanding of PCA, nor its relation to the course and cannot apply it at all to the images using a suitable software.	Demonstrates limited understanding of PCA and its relation to the course and can only apply it at a high level to the images using a suitable software.	Demonstrates adequate understanding of PCA and its relation to the course and can apply PCA with its steps to the images using a suitable software with limited understanding	Demonstrates a good understanding of PCA and its relation to the course and can apply PCA with its steps to the images using a suitable software with reasonable understanding	Demonstrates a good understanding of PCA and its relation to the course and can apply PCA with its steps to the images using a suitable software with complete understanding	
3.	**Perform error analysis,** demonstrate usefulness of PCA approach in comparison to its non-usage and **recommend a suitable number of Principal Components** to be used based on the error analysis. **(Marks=2)**	Cannot perform error analysis for changing number of principal components and thus cannot demonstrate the usefulness of the PCA approach nor can recommend a suitable number of Principal Components to be used.	Can only perform limited error analysis for changing number of principal components but cannot demonstrate the usefulness of the PCA approach nor can recommend a suitable number of Principal Components to be used.	Can perform adequate error analysis for changing number of principal components but demonstration of usefulness of the PCA approach and recommendation of suitable number of Principal Components is not adequate.	Can perform complete error analysis for changing number of principal components and demonstration of usefulness of the PCA approach and recommendation of suitable number of Principal Components is adequate but not complete.	Can perform complete error analysis for changing number of principal components and demonstration of usefulness of the PCA approach and recommendation of suitable number of Principal Components is complete.	

4.	CEP **Report (Marks=1)**	CEP report is not of a suitable format, and does not contain the important sections, and is full of spelling and grammar mistakes	CEP report is of a suitable format but does not contain the important sections, has many of spelling and grammar mistakes	CEP report is of a suitable format, contains important sections but with limited details, and has only some spelling and grammar mistakes	CEP report is of a suitable format, contains important sections with reasonable details, and has only few spelling and grammar mistakes	CEP report is of a suitable format, contains important sections with complete details, and has no spelling and grammar mistakes	
5,	**Task distribution** and Teamwork **(Marks=1)**	No mention of task distribution and teamwork	Inadequate mention of task distribution and teamwork	Adequate but limited mention of task distribution and teamwork but cannot justify it.	Reasonable details about task distribution and teamwork and can justify it to some extent.	Reasonable details about task distribution and teamwork and can completely justify it.	