Classification of historic buildings in Indonesia with svd algorithm based on texture and color characteristics (case: temples, historic mosques, historic churches, and historic Confucian temples)

Hadiyanto*, Manatap Dolok Lauro and Dyah Erny Herwindiati
Informatic Engineering Department, Faculty of Information Technology
Universitas Tarumanagara Jakarta Indonesia 11440

* hadiyyanto.535160107@stu.un.tar.ac.id

Abstract. This study aims to make an application that able to make classification for historical building in Indonesia. The building types being used in this study are: temples, historic mosques, historic churches and historic confucian temples. This application classifies historical buildings based on texture, Color and combination of two. GLCM method is utilzied for texture characteristics classification. Histogram method is being used for Color characteristics. SVD method is used to reduce dimensions, which can make classification process becomes faster.

1. Introduction
In Indonesia, historical buildings are utilized as a means of tourism because they are considered to have high historical value and charming building architecture. Especially the types of historic buildings such as temples, historic mosques, historic churches, and historic confucian temple which have a high tourist attraction because of their beauty but there are still many people who have difficulty in knowing the type of historic buildings they are see [1]. Based on previous problems, an application is develop to classify historic buildings in Indonesia (case: temples, historic mosques, historic churches and historic confucian temples).

2. Method and materials
2.1. Sample preparation
In this design only historic buildings are classified. Historical building image data obtained from the internet. The image data used has a size of 256x256 pixels and has a .jpg format. The historic buildings in question only consist of temples, historic mosques, historic churches and historic confucian temples in Indonesia.

![Figure 1. Sample image data of historic buildings in Indonesia](image)

2.2. Method
In this application the SVD method is used to help the process of reducing dimensions so that the classification process can be faster, GLCM is used to assist in the classification process
based on texture, the Histogram is used to assist in the classification process based on color and Mahalanobis Distance are used to find the distances.

2.2.1. SVD

Singular Value Decomposition (SVD) method is a matrix decomposition method that is broken down into multiplications of three matrices, each of which has certain characteristics namely the left singular vector matrix, the singular value matrix, and the right singular vector matrix. The SVD method can be seen as a method for reducing dimensions. In this design, the SVD method plays a role in reducing dimensions so that the classification process can be faster [2].

Steps of the SVD method:
1. Initialization of matrix variables. For example: A \([m, n]\): matrix A size \(m \times n\).
2. Transpose the matrix A.
3. Finding the value of \(M\) using the method \(M = A^T \times A\).
4. Finding the value of the eigen vector and eigen value by using the \(M\) value that has been obtained.
5. After the eigen vector and eigen value values have been obtained, data is sorted from large to small on the eigen vector value and the eigen value. Thus forming a new eigenvector and a new eigenvalue.
6. The cumulative ratio calculation process is performed to get \(V_{\text{choose}}\).
7. After getting the \(V_{\text{choose}}\) value, the next thing to do is to carry out the process of dimension reduction by using the method: \(A_{\text{new}} = A \times V_{\text{choose}}\).
8. Finally, carried out the process of transformation back from \(A_{\text{new}}\) to the original dimension, using the method \(A_{\text{reverse}} = A_{\text{new}} \times V_{\text{choose}}^T\).

2.2.2. GLCM

The Gray Level Co-occurrence Matrix (GLCM) method is one of the second extraction sequences on the texture statistical features. The GLCM method uses grayscale images. In this design the GLCM texture features that will be used are contrast, homogeneity, energy and correlation [3],[4],[5].

1. Contrast

Contrast is a measure of intensity or variation in gray level between pixels and their neighbors. Contrast formula:

\[
\text{Contrast} = \sum_{i=1}^{L} \sum_{j=1}^{L} (i - j)^2 \times \text{GLCM}(i, j)
\]  

(GLCM = GLCM matrix  
i = line to i  
j = column to j)

2. Homogeneity

Homogeneity measures how close the distribution of elements in the GLCM is to the diagonal of the GLCM. Homogeneity formula:

\[
\text{Homogeneity} = \sum_{(i,j)=1}^{L} \frac{\text{GLCM}}{1 + (i-j)^2}
\]  


3. Energy

Energy comes from Angular Second Moment (ASM). ASM measures local uniformity of gray levels. When the pixels are very similar, the ASM value will be large. Energy formula:

\[ ASM = \sum_{i=1}^{L} \sum_{j=1}^{L} (GLCM(i,j))^2 \]

\[ Energy = \sqrt{ASM} \]

4. Correlation

Correlation in GLCM measures the linear dependence of gray levels in the neighboring pixel images. Correlation formula:

\[ Correlation = \sum_{i=1}^{L} \sum_{j=1}^{L} \frac{(i-\mu_i)(j-\mu_j)GLCM(i,j)}{\sigma_i^2 \sigma_j^2} \]

2.2.3. Histogram

Histogram method is a method that distributes the intensity values of pixels in each color in a digital image in graphical form. In a digital image, the Histogram method can explain the number of pixels in each color by having a certain range of values that includes the color space of the digital image [6],[7]. Step of the Histogram method:

1. Divide the image based on the type of RGB color space.
2. Determine the number of bin using the formula:

\[ Bin[I] = X_{1\cdot Range} \cdots X_{1\cdot Range + Index} \]

Information:
I = bin to I

\[ Range = \frac{256}{bin} \]

\[ Index = 0 \cdot s/d \left( \frac{256}{bin} - 1 \right) \]
3. Calculates the frequency of many pixels according to a predetermined bin.
4. The value on the histogram is formed into a vector for each color. Then, the three vectors are combined into one which is the result of the histogram.

2.2.4. Mahalanobis Distance
Mahalanobis distance is a statistical method used to obtain data with a certain distance. A pixel vector value entered in one class is determined by a number of nearby samples [8]. Mahalanobis formula:

\[ d(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T S^{-1} (\mathbf{x} - \mathbf{y}) \]  

Information:
\( S^{-1} = \text{Covariance Matrix} \)

3. Results and discussion
3.1. Test results on 100 data images that have been trained
Testing of 100 image data that has been trained will be carried out using texture characteristics, color characteristics and characteristics of a combination of texture and color so that the percentage of success can be known. The results of 100 data images that have been trained:

| Experiment | Image Data | Image Category | Process Results | Classification Results |
|------------|------------|----------------|-----------------|------------------------|
| 1          | Temple     | Temple         | 1               | 0                      |
| 2          | Temple     | Temple         | 1               | 0                      |
| 3          | Temple     | Temple         | 1               | 0                      |

**Figure 2.** Sample testing of image data

**Table 1.** The results of testing 100 data images that has been trained

| The Characteristics Used | Percentage Of Success |
|--------------------------|------------------------|
| Texture                  | 90%                    |
| Color                    | 91%                    |
| Texture+Color            | 94%                    |
3.2. Test results on 120 data images that have been trained
Testing of 120 image data that has been trained will be carried out using texture characteristics, color characteristics and characteristics of a combination of texture and color so that the percentage of success can be known. The results of 120 data images that have been trained:

| The Characteristics Used | Percentage Of Success |
|--------------------------|-----------------------|
| Texture                  | 89.16%                |
| Color                    | 90.83%                |
| Texture+Color            | 93.33%                |

**Table 2.** The results of testing 120 data images that have been trained

3.3. Test results on 40 data images that have never been trained
Testing of 40 image data that have never been trained will be carried out using texture characteristics, color characteristics and characteristics of a combination of texture and color so that the percentage of success can be known. The results of 40 data images that have never been trained:

| The Characteristics Used | Percentage Of Success |
|--------------------------|-----------------------|
| Texture                  | 75%                   |
| Color                    | 77.5%                 |
| Texture+Color            | 85%                   |

**Table 3.** The results of testing 40 data images that have never been trained

3.4. Test results on 50 data images that have never been trained
Testing of 50 image data that have never been trained will be carried out using texture characteristics, color characteristics and characteristics of a combination of texture and color so that the percentage of success can be known. The results of 50 data images that have never been trained:

| The Characteristics Used | Percentage Of Success |
|--------------------------|-----------------------|
| Texture                  | 78%                   |
| Color                    | 82%                   |
| Texture+Color            | 84%                   |

**Table 4.** The results of testing 50 data images that have never been trained

4. Conclusion
According to test result found in table 1, 2, 3, and 4, the combined characteristics of texture and color proven to be the best method. The percentage of success obtained at the time of testing the data using the combined texture and color characteristics get a percentage of success the biggest compared to the percentage of success using texture and color characteristics.
5. References

[1] Ana Prasetyowati, 2008 *Perlindungan Karya Cipta Bangunan Kuno/Bersejarah Di Kota Semarang Sebagai Warisan Budaya Bangsa*.

[2] Geraldi Lukito Nagaria, Dyah Erny Herwindiati, Sidik Mulyono, 2016 *Klasifikasi Lahan Mangrove Menggunakan Metode Singular Value Decomposition*.

[3] Hafid Satrio Priambodo, 2019 *Klasifikasi Jenis Citra Makanan menggunakan Color Histogram dan Gray Level Co-Occurrence Matrix dengan K-Nearest Neighbour* vol. 3.

[4] Ezra Alvansga, 2019 *Pengenalan Tekstur Menggunakan Metode GLCM Serta Modul Nirkabel*.

[5] Kadek Novar Setiawan, I Made Suwija Putra, 2018 *Klasifikasi Citra Mammogram Menggunakan Metode K-Means, GLCM, dan Support Vector Machine (SVM)* vol. 6.

[6] Kadir 2013 *Teori dan Aplikasi Pengolahan Citra* (Yogyakarta: Andi).

[7] Sangoh Jeong 2001 *Histogram-Based Color Image Retrieval* (Standford University).

[8] Achmad Rizal, 2008 *Pengenalan Identitas Manusia Melalui Pola Iris Mata Menggunakan Transformasi Wavelet dan Mahalanobis Distance*.