

## CLASSIFICATION OF BATIK LAMONGAN BASED ON FEATURES OF COLOR, TEXTURE AND SHAPE

<sup>a</sup>Miftahus Sholihin, <sup>b</sup>Siti Mujilahwati, <sup>c</sup>Retno Wardhani

<sup>a,b,c</sup>Fakultas Teknik, Universitas Islam Lamongan

E-Mail: miftahus.sholihin@unisla.ac.id

### *Abstract*

*Batik is as one of Indonesian cultural heritage, which has grown almost throughout the regions of Indonesia. Its very natural when each region has a patterns and motifs of different batik. The One of batik motif quite famous is batik from Lamongan (Sendang Batik). Classification aims to classify object into specific classes based on the value of the attribute associated with the object being observed. In this research designed a system that serves to classify Lamongan batik cloth based on color features using color moment, texture using Gray Level Co-occurrence Matrix (GLCM), and shape using moment invariant, classification using K-Nearest Neighbors (K-NN) method. The classification of batik images is based on three classes namely slempang, pethetan, and putihan class. The amount of image data used is 120 which divided into training data and testing data. In outline the system was built consists of three main processes namely pre-processing, feature extraction, and classification. The highest accuracy rate in this study was 90.4% when the value of  $k = 6$ .*

*Keywords: Batik, Classification, Preprocessing, Feature Extraction, Gray Level Co-occurrence Matrix, K-Nearest Neighbors*

## **INTRODUCTION**

Batik is one of the works of ancestors of the Indonesian nation that has been recognized by UNESCO as an international cultural heritage on October 2, 2009. The development of batik is not only on the island of Java. However, batik spread and expanded throughout Indonesia. The development of batik produces new motifs and patterns in each region.

An image can be visually recognized based on its features. Selection of the right features can provide detailed information about the class of an image. Some features that can be extracted from an image are color, texture, and shape. Texture analysis is one of the techniques of image analysis based on the assumption that the image is formed by the variation of pixel intensity, both grayscale and color image [1]. Texture analysis is done in an effort to extract the features or characteristics of an image in order to be able to recognize or differentiate the image in a class with an image in another class.

Several research have been done to extract the image of batik. Nurhaida et al [2] in his research stated that gray level co-occurrence matrix (GLCM) is the best features extraction method when compared with canny edge detector and gabor filter to recognize batik motif. Siqueira et al [3] stated that among several statistical approaches, GLCM has proven to be very powerful as feature descriptor or features in representing texture features of an image. The GLCM method only uses eight features is capable of achieving the highest classification accuracy when compared with the Gabor filter method, Discrete Wavelet Transform (DWT) (3 features) 60.90%, Granulometrics (20 features) 91.13%, and Local Binary Pattern (LBP) (10 features) 89.51% [4], [5].

Nugroho et al [6] classified batik by comparing the extraction of invariant moment, eccentricity, and compactness features. In this research the classification is divided into five classes based on the origin of the region. From the research, it is found that feature extraction with moment invariant gives better result than eccentricity and compactness. Kurniawardhani et al [7] classified using the method of feature

extraction invariant to rotation. From the research that has been done the maximum accuracy obtained by 90%.

Rangkuti et al [8] conducted research on content based batik image retrieval. The research is done using the features of texture by using wavelet and features shape by using moment invariant. The results of the study provide an average of 90-92% accuracy. Another research conducted [9] which carried out the classification of batik images, where the feature extraction method using GLCM and geometric moment invariant. Classification with geometric moment invariant features gives 80% accuracy, while the accuracy using GLCM is 70%.

Rangkuti et al [10] in his research used texture and shape features. The texture feature used is Daubechies wavelet, while the shape feature used invariant moments. The highest accuracy of this research is 90%.

Wahyuningrum et al [11] in her research said the use of a combination of two features gives better results when compared using only one features. Patil et al [12] in his research stated the use of a combination of color, texture, and shape features provides much better results when compared only used one feature.

Based on the description above, then in this research will design a system that serves to perform Lamongan batik classification based on texture, color, and shape features. The features extraction method used is GLCM, color moment, and moment invariant. While the classification using KNN.

## **CLASSIFICATION OF BATIK IMAGE**

Data was used in this research is Lamongan batik image. The amount of image data was used in the research is 120 images that are divided into two, namely image for training data and testing data.

Figure 1 is a general overview of the system created. Based on Figure 1 there are three main processes: pre-processing, feature extraction, and classification.

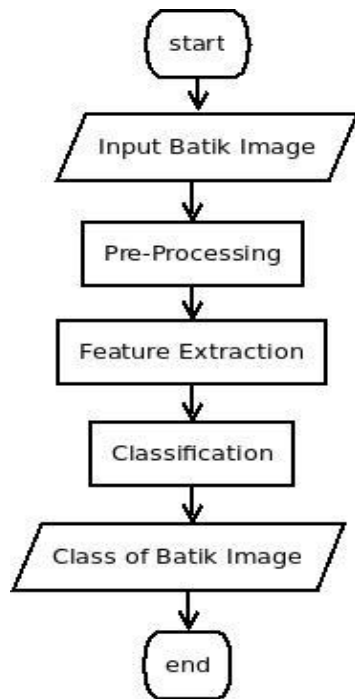


Figure 1. General overview of the system

### Preprocessing

The purpose of preprocessing is to improve image quality. In this research pre-processing was conducted is to change the image size from the initial size to 256x256, in addition to pre-processing is to change the image from Red, Green, Blue (RGB) to grayscale, RGB to Hue, Saturation, Value (HSV).

### Feature Extraction

In this research using three methods of feature extraction that color moment features extraction, texture features extraction, and shape features extraction.

### Color Feature Extraction

Color moments are a method was used to distinguish images based on their color feature. This method assumes that the color distribution in an image can be expressed as a probability distribution. The basis of color moments is the color distribution in the image can be interpreted as the probability of the distribution. If the color in the image follows the probability of a particular distribution, then the moment of the distribution can be used as a feature for the identification of images by color [13].

Some color distribution information is arranged in three sequences of moment. The first moment ( $\mu$ ) is the average pixel value ( $P_{ij}$ ) in each color channel [14]. If  $P_{ij}$  is a pixel  $j$  in the color channel  $i$  and  $N$  is the sum of all image pixels, then the first moment ( $\mu$ ) can be calculated by the equation (1). The second moment ( $\sigma$ ) described the standard deviation shown in Equation (2), and the next moment ( $\theta$ ) represents the inclination of the color shown in Equation (3) [15]. Figure 2 is a color moment process.

Jose et al [16] in his research stated that the HSV method is the best method of color recognition. The HSV color space provides better results when compared to RGB and YCbCr [17], [18].

$$\mu_c = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N p_{ij}^c \quad (1)$$

$$\sigma_c = \left[ \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (p_{ij}^c - \mu_c)^2 \right]^{\frac{1}{2}} \quad (2)$$

$$\theta_c = \left[ \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (p_{ij}^c - \mu_c)^3 \right]^{\frac{1}{3}} \quad (3)$$

### Texture Feature Extraction

GLCM is one of the methods for obtaining a second-order statistical feature by calculating the probability of an adjacency relationship between two pixels at a certain distance and angle orientation [19], [20].

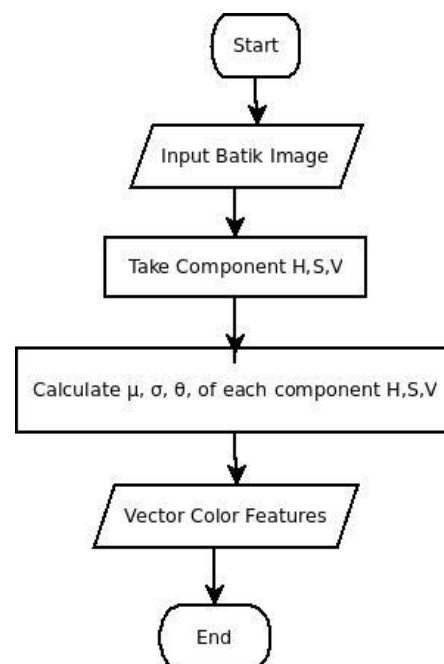


Figure 2. Color Moment Feature Extraction

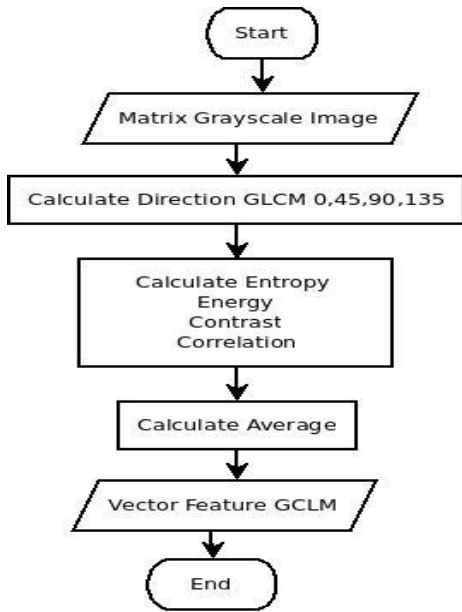


Figure 3. Texture Feature Extraction Process

Figure 3 is a feature texture extraction process. The first step to calculate the GLCM feature is to convert the RGB image into a gray scale image. The second step is to create a co-occurrence matrix and proceed by determining the spatial relationship between the neighboring pixels based on angle  $\theta$  and distance  $d$ . The next step is to create a symmetric matrix by adding a co-occurrence matrix with its transpose matrix. Then normalization of the symmetric matrix by calculating the probability of each matrix. The final step is to calculate the GLCM feature. Each feature is calculated by one pixel distance in four directions, ie  $0^0$ ,  $45^0$ ,  $90^0$ , and  $135^0$  to detect co-occurrence.

When GLCM has a matrix with a size  $L \times L$ , in which  $L$  is the number of gray levels of the original image and when the probability of pixel  $i$  is the neighbor of pixel  $j$  within distance  $d$  and edge orientation  $\theta$  is  $P$ , the energy feature, the entropy feature, the contrast feature and the correlation feature can be calculated by Equations (4), (5), (6), and (7).

$$ASM = \sum_{i,j=0}^{L-1} p^2(i, j, d, \theta) \quad (4)$$

$$Ent = \sum_{i,j=0}^{L-1} p^2(i, j, d, \theta) \cdot \log p(i, j, d, \theta) \quad (5)$$

$$Cont = \sum_{i,j=0}^{L-1} (i - j)^2 \cdot p(i, j, d, \theta) \quad (6)$$

$$Corr = \sum_{i,j=0}^{L-1} \frac{(i - \mu_x)(j - \mu_y) \cdot p(i, j, d, \theta)}{\sigma_x \sigma_y} \quad (7)$$

Wherein,

$$\mu_x = \sum_{i,j=0}^{L-1} i \cdot P(i, j, d, \theta),$$

$$\mu_y = \sum_{i,j=0}^{L-1} j \cdot P(i, j, d, \theta),$$

$$\sigma_x = \sum_{i,j=0}^{L-1} (i - \mu_x)^2 \cdot P(i, j, d, \theta),$$

$$\sigma_y = \sum_{i,j=0}^{L-1} (j - \mu_y)^2 \cdot P(i, j, d, \theta)$$

### Shape Feature Extraction

Moments (moment) can describe an object in terms of area, position, orientation and other undefined parameters. The function of continuous 2 dimension  $f(x, y)$ , moment orde  $(p+q)$  is defined by Equation (8) [21].

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad (8)$$

wherein  $p, q = 0, 1, 2$ .

This moment depend on size, translation, and rotation. Moment independent of translation and rotation are expressed by a central moment defined by Equation (9).

$$\mu_{pq} = \sum_{x=1}^M \sum_{y=1}^N (x - \hat{x})^p (y - \hat{y})^q f(x, y) \quad (9)$$

wherein  $\hat{x} = \frac{m_{10}}{m_{00}}$ , and  $\hat{y} = \frac{m_{01}}{m_{00}}$ .

Normalized central moments are defined by Equation (10).

$$\eta_{pq} = \frac{\mu_{pq}}{(\mu_{00})^\lambda} \quad (10)$$

wherein  $\lambda = \frac{(i+j)}{2} + 1$

Based on the normalized moment, Hu [22] introduces seven invariants according to Equations (11), (12), (13), (14), (15), (16), and (17).

$$\phi_1 = \eta_{20} + \eta_{02} \quad (11)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad (12)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (13)$$

$$\phi_4 = (\eta_{30} - 3\eta_{12})^2 + (\eta_{21} - \eta_{03})^2 \quad (14)$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \quad (15)$$

$$[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + 3(\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) + [3(\eta_{30} + \eta_{21})^2 - (\eta_{21} + \eta_{03})^2]$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{21}^2)] + 4\eta_{11}(\eta_{30} + \eta_{12}) + (\eta_{21} + \eta_{03}) \quad (16)$$

$$\phi_7 = \frac{(3\eta_{21} - \eta_{03})(\eta_{30} - \eta_{12})}{[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{21} + \eta_{03})^2]} \quad (17)$$

$$\frac{(3\eta_{21} + \eta_{30})(\eta_{21} + \eta_{03})}{[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03}^2)]}$$

Figure 4 is an overview of the shape feature extraction process.

**Classification**

Classification aims to classify object into specific classes based on the value of the attributes associated with the object being observed. Each particular object has certain features, thus the classification can distinguish an object from another object.

In this research the classification method used is K-Nearest Neighbor (KNN). Gu and Song [23] in his research say that KNN achieves higher accuracy results than other classification algorithms. The purpose of this algorithm is to classify new objects based on attributes and training data. This algorithm works based on the minimum distance of new data to the nearest neighboring *k*. This method is the main thing to do is to determine the number of *k* that will be used for the classification process.

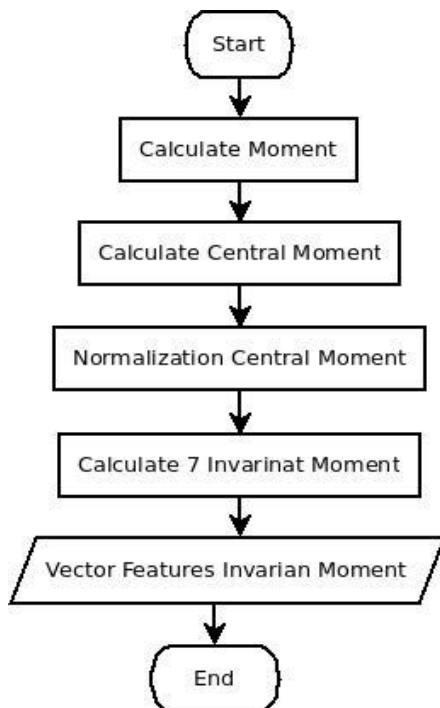


Figure 4. Shape Feature Extraction

**RESULT AND DISCUSSION**

Trial was conducted from this research aimed to determine the class of batik image entered by the user. Batik image will be classified into three classes namely slempang class, pethetan, and putihan. The amount of data used in this research is 120 which is divided into two of 35 images as training data and 85 images used as data testing. The image size was used in the research is 256x256. Features used amounted to 21 features consisting of 9 color features, 4 texture features, and 7 features of the shape.

In this research the classification process is divided into two stages.

**Training Process**

This process aimed at training the system. The training process was done only one time. The system has not been able to provide conclusion or result when the training has not been done. Figure 5 is a training process. Training process begins by taking the image to be entered into the system. The next process is to do preprocessing. The image that has been done preprocessing process, the next process is to extract the features of the image. The last process is to save the image feature into the database or into a file. Features data stored into the database that will be used as data knowledge to determine the class of the input image.

Data used for training process is not used anymore for the testing process.

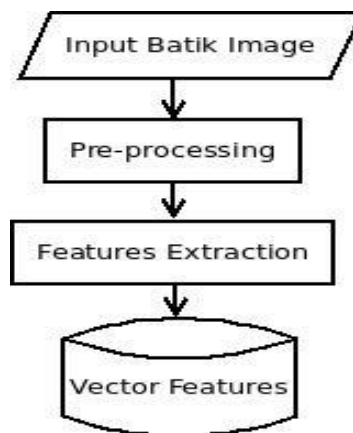


Figure 5. Training Process

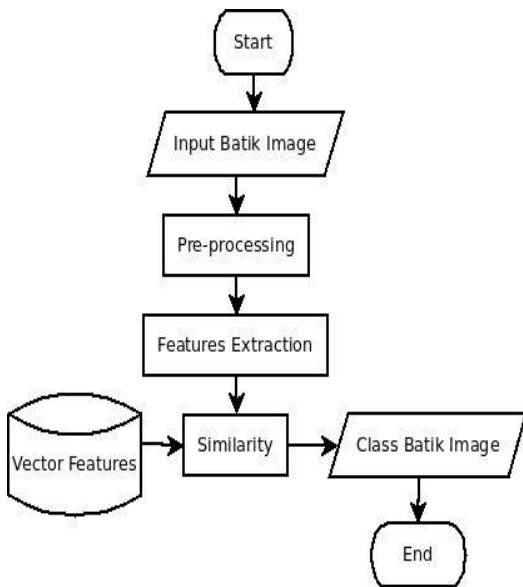


Figure 6. Testing Process

**Testing Process**

The stages in the testing process are not much different from those in the training process. In the testing process the system will give the results of the input provided by the user. Based on the image of the input system will classify the image data based on training data that has been done. Figure 6 is a testing process.

Based on the test results shown Table 1 obtained the highest accuracy of 90.4% when the value  $k = 6$ . Based on experiment that has been done, the system is still not able to classify data in accordance with the class of batik images correctly. The system is able to recognize the class of putihan perfectly, this is because the image in this putihan class has a simple motif, the basic color of the image very different from the petethan class and also the slempang class. While for pethetan class and slempang system give result which not very perfect, this is because base color of this motif many resemblance between slempang with pethetan. In addition, the texture and color also have similarities, so the results cannot be maximized classification.

**CONCLUSION**

Based on the experiment that have been done, can be drawn some conclusions, among others:

1. Maximum accuracy gained 90.4% when the value of neighborhood or  $k = 6$ .
2. The class of Putihan gives maximum results for the process of classification.

Table 1. Experimental Result

K	Class	Result		
		True	False	Accuracy
1	Slempang	16	4	78,8%
	Petethan	40	14	
	Putihan	11	0	
2	Slempang	16	4	78,8%
	Petethan	40	14	
	Putihan	11	0	
3	Slempang	16	4	78,8%
	Petethan	40	14	
	Putihan	11	0	
4	Slempang	16	4	78,8%
	Petethan	40	14	
	Putihan	11	0	
5	Slempang	19	1	89,4%
	Petethan	46	8	
	Putihan	11	0	
6	Slempang	19	1	90,5%
	Petethan	47	7	
	Putihan	11	0	
7	Slempang	20	0	89,4%
	Petethan	45	9	
	Putihan	11	0	
8	Slempang	16	4	89,4%
	Petethan	45	9	
	Putihan	11	0	
9	Slempang	16	4	89,4%
	Petethan	45	9	
	Putihan	11	0	
10	Slempang	16	4	89,4%
	Petethan	45	9	
	Putihan	11	0	

## REFERENCES

- [1] H. Wibawanto and A. Susanto, "Analisis Tekstur Untuk Diskriminasi Massa Kistik Dan Non-Kistik Pada Citra Ultrasonografi," Universitas Gadjah Mada, 2011.
- [2] I. Nurhaida, R. Manurung, and A. M. Arymurthy, "Performance Comparison Analysis Features Extraction Methods for Batik Recognition," in *Advanced Computer Science and Information Systems (ICACSIS), 2012 International Conference on*, 2012, pp. 207–212.
- [3] F. R. de Siqueira, W. R. Schwartz, and H. Pedrini, "Multi-Scale Gray Level Co-Occurrence Matrices for Texture Description," *Neurocomputing*, vol. 120, pp. 336–345, 2013.
- [4] F. Mirzapour and H. Ghassemian, "Using GLCM and Gabor Filters for Classification of PAN Images," in *Electrical Engineering (ICEE), 2013 21st Iranian Conference on*, 2013, pp. 1–6.
- [5] L. A. Ella, F. van den Bergh, B. J. van Wyk, and M. A. van Wyk, "A Comparison of Texture Feature Algorithms for Urban Settlement Classification," in *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, 2008, vol. 3, p. III-1308-III-1311.
- [6] N. A. Haryono, W. Hapsari, A. Angesti, and S. Felixiana, "Penggunaan Momen Invariant, Eccentricity, dan Compactness Untuk Klasifikasi Motif Batik Dengan K-Nearest Neighbour," *Inform. J. Teknol. Komput. Dan Inform.*, vol. 11, no. 2, 2016.
- [7] A. Kurniawardhani, N. Suciati, and I. Ariesanti, "Klasifikasi Citra Batik Menggunakan Metode Ekstraksi Ciri yang Invariant Terhadap Rotasi," *JUTI J. Ilm. Teknol. Inf.*, vol. 12, no. 2, pp. 48–60, 2014.
- [8] H. Rangkuti, A. Harjoko, and A. E. Putro, "Content Based Batik Image Retrieval," 2014.
- [9] N. L. W. S. R. Ginantra, "Deteksi Batik Parang Menggunakan Fitur Co-Occurrence Matrix dan Geometric Moment Invariant Dengan Klasifikasi KNN," *Lontar Komput. J. Ilm. Teknol. Inf.*, pp. 40–50.
- [10] A. H. Rangkuti, Z. E. Rasjid, and D. J. Santoso, "Batik Image Classification Using Treeval and Treefit as Decision Tree Function in Optimizing Content Based Batik Image Retrieval," *Procedia Comput. Sci.*, vol. 59, pp. 577–583, 2015.
- [11] R. T. Wahyuningrum and I. A. Siradjuddin, "An Efficient Batik Image Retrieval System Based On Color And Texture Features," *J. Theor. Appl. Inf. Technol.*, vol. 81, no. 2, p. 349, 2015.
- [12] J. K. Patil and R. Kumar, "Analysis of Content Based Image Retrieval for Plant Leaf Diseases Using Color, Shape and Texture Features," *Eng. Agric. Environ. Food*, vol. 10, no. 2, pp. 69–78, 2017.
- [13] Y. Herdiyeni and M. M. Santoni, "Combination of Morphological, Local Binary Pattern Variance and Color Moments Features for Indonesian Medicinal Plants Identification," in *Advanced Computer Science and Information Systems (ICACSIS), 2012 International Conference on*, 2012, pp. 255–259.
- [14] S. R. Kodituwakku and S. Selvarajah, "Comparison of Color Features for Image Retrieval," *Indian J. Comput. Sci. Eng.*, vol. 1, no. 3, pp. 207–211, 2004.
- [15] W. L. Martinez and A. R. Martinez, *Computational statistics handbook with MATLAB*, vol. 22. CRC press, 2007.
- [16] J. M. Chaves-González, M. A. Vega-Rodríguez, J. A. Gómez-Pulido, and J. M. Sánchez-Pérez, "Detecting Skin in Face Recognition Systems: A Colour Spaces Study," *Digit. Signal Process.*, vol. 20, no. 3, pp. 806–823, 2010.

- [17]J. K. Patil and R. Kumar, "Feature Extraction & Retrieval of Plant Leaf Diseases Using Image Histogram," *Int J Electron Comput*, vol. 5, no. 2, pp. 143–151, 2013.
- [18]J. K. Patil and R. Kumar, "Plant Leaf Disease Image Retrieval Using Color Moments," *IAES Int. J. Artif. Intell.*, vol. 2, no. 1, 2013.
- [19]F. Albrechtsen, "Statistical Texture Measures Computed from Gray Level Cooccurrence Matrices," *Image Process. Lab. Dep. Inform. Univ. Oslo*, vol. 5, 2008.
- [20]D. Gadkari, "Image Quality Analysis Using GLCM," 2004.
- [21]R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Prentice hall New Jersey, 2008.
- [22]M.-K. Hu, "Pattern Recognition by Method of Moments," *IRE Trans. Inf. Theory*, vol. 8, no. 2, pp. 179–187, 1961.
- [23]Q. Gu and Z. Song, "Image Classification Using SVM, KNN and Performance Comparison with Logistic Regression," *CS44 Final Proj. Rep.*, 2009.