Classification of stamps and handmade batik based on pattern recognition

M D Firdaus¹* and H Nugroho²

1 Department of Informatics, UIN Sunan Gunung Djati Bandung, Jalan A H Nasution 105 Bandung 40614, Indonesia
2 Telkom University, Jl. Telekomunikasi Terusan Buah Batu Indonesia 40257, Bandung, Indonesia

*deden@uinsgd.ac.id

Abstract. Batik is a traditional cloth made by painting motifs with Indonesian culture nuance on plain cloth. It has always been used by the people, from their birth until marriage, even someone who died using batik cloth as a cover. Batik cloth made with philosophy and symbolic meaning. Every region in Indonesia has a characteristic according to the circumstances and character of the man himself. There are three ways to make batik, such as using stamps, handmade, printing. Every type of batik has different types of patterns, and there is a repetition of a pattern and a combination of several patterns. Based on the type of batik, handmade batik has the highest price, and printing batik has the lowest prices. For batik lovers, handmade batik would want to have, but not everyone knows how to characterize the stamps batik and often mistaken in the detection of the type of batik. Thus, only a few people can distinguish the differences between the three types of batik. It is difficult for ordinary people to recognize a batik pattern. The purpose of this research is to create a prototype system that can detect the patterns of batik motifs created by handmade or stamps. Cosine similarity is used in this research to recognize the pattern and classified handmade or stamps batik. The results of this research is 95% in detecting handmade and stamp batik, then 91% in detecting sum of pattern in stamps batik.

1. Introduction
Batik is a traditional cloth made by painting motifs with Indonesian’s culture nuance on plain cloth. Batik is one of the Indonesian cultural products belonging to traditional clothing [1]. There are two kinds of batik, that is handmade and stamps batik. The study of batik has been done by researchers, especially in detecting batik motive. Handmade batik usually has regularly looped in patterns but just a little, difference with stamps batik which has so many regular looping in patterns, based on batik expert said that stamps batik has at least eight regularly looping in patterns.
Because kinds of batik motive are so many, for examples batik parang, lunglungan, cirebon, kawung, sekar, and etc., there are some research in detecting batik motive. One of research is Automatic batik motifs classification using various combination of SIFT Features Moments and k-Nearest Neighbor [2], this research tries to develop method to make classifications for batik parang, lereng, dutch, chinese, ceplokan, semen and lunglungan, and the result of accuracy is 31.43% [2]. The other research about detecting batik motifs is Batik Motif Classification using Color-Texture-Based feature extraction and backpropagation neural network [3]. In this research the motifs used to detecting are batik parang, ceplokan, lereng, mega mendung, semen, lunglungan and buketan [3]. The experiment of this research show that the software was developed can recognize batik motif quite well, with rate of Tanimoto distance 0.37. The study of batik classification was also carried out by Teny Handayani, where classification was carried out on several types of motifs, namely friends, mount ringgit, krecek, and bledak [4]. For the study of the classification of batik motifs, Aris Fanani also used Cardinal Spline Curve Representation for the classification of batik motifs [5]. Agus Eko Minarno was used method Gray Level Co-Occurrence Matriks to classified batik motifs [6].

Previous research about batik are focusing to detecting kinds of motif, but for this research is different with previous research, this research is will developing method to classification between handwritten and stamps batik, and count how many pattern in stamps batik, so this research focus to classification kind in developing batik not to batik motifs. The research which is almost similar to this research is about the detection of the types of original batik and not original batik. This research categorizes original batik consisting of stamped batik and handwritten batik, while non-original batik is batik manufacturer [7].

The results of accuracy in this research is 83% in classification handwritten and stamps batik, and 66% in detecting sum of patterns in stamps batik.

2. Method

These works consist of three important phases. First, scan some templates in images, in this case the position of templates is four positions, that is left top, right top, bottom left, and right bottom. Second, scan all template to whole image area by per pixels. Third, analyze the results of a scan from 4 positions and make clustering.

![Figure 2. Block diagram of clustering handmade and stamp batik.](image)

Figure 2 illustrates the global process of clustering stamp and handmade batik design system. Before the images start processing, size image will be reducing. In crop templates will create cropped images with size n x n, where n is 1/8 length images. In this case crop position is left top, right top, left bottom, and right bottom. To get similarity between template with each position images use method Cosine.
Similarity. If the similarity is high the value will near to one and if the similarity is low the value will near to zero. All of value results scan will analyze to make a conclusion that is stamp batik or handmade batik, and how many patterns that image batik.

Figure 3 shows the design process detail of clustering handmade and stamps. In this process consist of nine processes.

![Figure 3](image)

**Figure 3.** Design process detail of clustering handmade and stamps batik.

### 2.1. Define kernel/template

In this process is creating template/kernel where the size of template/kernel is n x n where n is 1/8 of length images. Template/kernel create from four position, left top, right top, bottom top, bottom left.

![Figure 4](image)

**Figure 4.** Define template/kernel.

Figure 4 illustrates four positions of templates in batik images. The purpose of this process is to get section patterns from batik and next scan to the whole image.

### 2.2. Correlate image using cosine similarity

After defining templates/kernel then every kernel correlating to whole image area by per pixels.
Figure 5 illustrates scanning process templates to each area images by per pixels. In this process use cosine similarity to get level of similarity between template and scan positions.

In this process using cosine similarity. Cosine similarity is a commonly used similarity measure for real-valued vectors, used in (among other fields) information retrieval to score the similarity of documents in the vector space model. In machine learning, common kernel functions such as the RBF kernel can be viewed as similarity functions [8]. The cosine of 0° is 1, and it is less than 1 for any angle in the interval (0,π) radians. It is thus a judgment of orientation and not magnitude: two vectors with the same orientation have a cosine similarity of 1, two vectors oriented at 90° relative to each other have a similarity of 0, and two vectors diametrically opposed have a similarity of -1, independent of their magnitude. The cosine similarity is particularly used in positive space, where the outcome is neatly bounded in (0,1). The name derives from the term "direction cosine": in this case, unit vectors are maximally "similar" if they're parallel and maximally "dissimilar" if they're orthogonal (perpendicular). This is analogous to the cosine, which is unity (maximum value) when the segments subtend a zero angle and zero (uncorrelated) when the segments are perpendicular.

These bounds apply for any number of dimensions, and the cosine similarity is most commonly used in high-dimensional positive spaces. For example, in information retrieval and text mining, each term is notionally assigned a different dimension and a document is characterised by a vector where the value in each dimension corresponds to the number of times the term appears in the document. Cosine similarity then gives a useful measure of how similar two documents are likely to be in terms of their subject matter [9]. The technique is also used to measure cohesion within clusters in the field of data mining.

The term cosine distance is often used for the complement in positive space, that is:

$$D_c(A,B) = 1 - S_c(A,B)$$

where $D_c$ the cosine distance and $S_c$ is the cosine similarity. It is important to note, however, that this is not a proper distance metric as it does not have the triangle inequality property—or, more formally, the Schwarz inequality—and it violates the coincidence axiom; to repair the triangle inequality property while maintaining the same ordering, it is necessary to convert to angular distance (see below).

One advantage of cosine similarity is its low-complexity, especially for sparse vectors: only the non-zero dimensions need to be considered.

Other names of cosine similarity are Orchini similarity and the Tucker coefficient of congruence; Ochiai similarity (see below) is cosine similarity applied to binary data.

$$\text{Similarity} = \cos(\theta) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

(1)

Results of this process are store value to a matrix. Matrix results is four which references to four position templates.
Figure 6. Illustration of matrix result scan which show in diagram.

Figure 7. Merging into one matrix.

Figure 7show the result of merging matrix in four positions. It is shown connected component area. If sum of a connected component more than 24 for it classify to stamps batik and if less than 24 classify to handmade batik. After that, if image classifies to stamps batik then find to know how many patterns in that image.

2.3. Counting sum of pattern in stamps batik
In this process is how to calculate the sum of pattern in stamps batik, because stamps batik has regularly looping of pattern in their image. There are three feature is used to get sum of patterns, that is distances, angle and image moment.

2.3.1. Distances. In this process distance used to get distance from every centroid of nearest connected component. The distance can show regularly looping every pattern which have the same motifs.

2.3.2. Angle. Besides based on the closest distance between neighbors, an angle is also calculated between the point to the nearest neighbor, this is because there is a diagonal batik pattern so that the angle between the points is also likely to have the same angle.
2.3.3. Image moment. Image moment used to analyze the contour of every connected component. Moment is used in this process is spatial moment and center moment which invariant to translation. Below are the algorithm of spatial moment and center moment.

Image moment is a certain particular weighted average (moment) of the image pixels' intensities or a function of such moments, usually chosen to have some attractive property or interpretation [10].

Image moments are useful to describe objects after segmentation. Simple properties of the image which are found via image moments include area (or total intensity), its centroid, and information about its orientation.

For a 2D continuous function \( f(x,y) \) the moment (sometimes called "raw moment") of order \((p + q)\) is defined as

\[
M_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x,y) dx \, dy
\]

for \( p,q = 0,1,2,... \) Adapting this to scalar (greyscale) image with pixel intensities \( I(x,y) \), raw image moments \( M_{ij} \) are calculated by

\[
M_{ij} = \sum x^i \sum y^j I(x,y)
\]

In some cases, this may be calculated by considering the image as a probability density function, i.e., by dividing the above by

\[
\sum x \sum y I(x,y)
\]

A uniqueness theorem (Hu [1962]) states that if \( f(x,y) \) is piecewise continuous and has nonzero values only in a finite part of the \( xy \) plane, moments of all orders exist, and the moment sequence \( (M_{pq}) \) is uniquely determined by \( f(x,y) \). Conversely, \( (M_{pq}) \) uniquely determines \( f(x,y) \). In practice, the image is summarized with functions of a few lower order moments.

Simple image properties derived via raw moments include:

- Area (for binary images) or sum of grey level (for grey tone images): \( M_{00} \).
- Centroid : \( \{\bar{x}, \bar{y}\} = \frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}} \)

3. Results and discussion

For testing system use 200 images which contain 100 of stamps batik and 100 handmade batik. Below is result of detecting 200 images batik.

| Kind of batik | Detected Stamp | Detected Handmade | Accuracy |
|---------------|----------------|--------------------|----------|
| Stamp         | 100            | 0                  | 100%     |
| Handmade      | 10             | 90                 | 90%      |
|               |                |                    | **95%**  |

For result in detecting the sum of pattern in stamp batik is below.

| Images        | Accurate detected | Failure detected | Accuracy |
|---------------|-------------------|------------------|----------|
| 100 images of stamps batik | 91                | 9                | **91%**  |

From the table above the system has some failure detected in handmade batik. There is some reason to make these failures such us, image not scratch, lighting not standard. But the system can detect 100% accurate to stamps batik even in bad images. Results of this research are a difference with previous research because this research focus to detect handmade and stamps batik not focus on detecting kind of motifs batik and motifs batik for this research is unlimited.
4. Conclusion
From this research can make a conclusion that the system developed can classify between stamps and handmade batik in accuracy 95% from 200 images, and for detecting sum of pattern in stamps batik the accuracy is 91%.

References
[1] Nava N 1991 *II batik* (Ulissedizioni)
[2] Setyawwan I, Ivanna K, Timotius and Kalvin M 2015 *Automatic Batik Motifs Classification using Various Combinations of SIFT Features Moments and k-Nearest Neighbor (IEEE)*
[3] Suciati N, Pratomo W A and Purwitasari D 2014 Batik Motif Classification using Color-Texture-Based Feature Extraction and Backpropagation Neural Network (IEEE)
[4] Handhayani T 2016 Batik Lasem Images Classification Using Voting Feature Intervals 5 and Statistical Features Selection Approach (IEEE)
[5] Fanani A, Yuniarti A and Suciati N 2014 Geometric Features Extraction of Batik Images Using Cardinal Spline Curve Representation *Telkomnika Journal* 12(2)
[6] Minarno A E and Suciati N 2014 Batik Image Retrieval Based on Color Difference Histogram and Gray Level Co-Occurrence Matrik *Telkomnika Journal* 12(3)
[7] Yunari N and Yuniarno E M and Purnomo M H 2017 Indonesian Batik Image Classification Using Statistical Texture Feature Extraction Gray Level Co-occurrence Matrix (GLCM) and Learning Vector Quantization (LVQ) *Journal of Telecommunication, Electronic and Computer Engineering*
[8] Vert J P, Tsuda K and Schölkopf B 2004 A primer on kernel methods *Kernel Methods in Computational Biology*
[9] Singhal A 2001 Modern Information Retrieval: A Brief Overview *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* 24(4) 35–43
[10] Rangkuti A H, Bahaweres R B and Harjoko A 2012 Batik Image Retrieval Based on Similarity of Shape and Texture Characteristics (*ICAColis*)