

# Improving the Performance of CBIR on Islamic Women Apparels Using Normalized PHOG

Cut Mutia<sup>1</sup>, Fitri Arnia<sup>\*2</sup>, Rusdha Muharar<sup>3</sup>

<sup>1,2,3</sup> Graduate Program in Electrical Engineering, Syiah Kuala University, Banda Aceh, Indonesia

<sup>2,3</sup> Department of Electrical and Computer Engineering, Syiah Kuala University, Banda Aceh, Indonesia

<sup>\*</sup>corresponding author, e-mail: mutiacut50@yahoo.co.id<sup>1</sup>, f.arnia@unsyiah.ac.id<sup>2</sup>,

r.muhaarar@unsyiah.ac.id<sup>3</sup>

## Abstract

*The designs of Islamic women apparels is dynamically changing, which can be shown by emerging of online shops selling clothing with fast updates of newest models. Traditionally, buying the clothes online can be done by querying the keywords to the retrieval system. The approach has a drawback that the keywords cannot describe the clothes designs precisely. Therefore, a searching based on content—known as content-based image retrieval (CBIR)—is required. One of the features used in CBIR is the shape. This article presents a new normalization approach to the Pyramid Histogram of Oriented Gradients (PHOG) as a mean for shape feature extraction of women Islamic clothing in a retrieval system. We refer to the proposed approach as normalized PHOG (NPHOG). The Euclidean distance measured the similarity of the clothing. The performance of the system was evaluated by using 340 clothing images, comprised of four clothing categories, 85 images for each category: blouse-pants, long dress, outerwear, and tunic. The recall and precision parameters measured the retrieval performance; the Histogram of Oriented Gradients (HOG) and PHOG were the methods for comparison. The experiments showed that NPHOG improved the HOG and PHOG performance in three clothing categories.*

**Keywords:** Muslimah fashion, CBIR, shape feature, NPHOG, Euclidean Distance

## 1. Introduction

Today, Islamic women fashion is becoming more popular, which is indicated by the increased number of market demand on various styles of Muslimah clothing. The latest models keep appearing every month even in a matter of weeks. The development of information technology also affects the way the people shops; the sales of Muslimah clothes do not take place only in the traditional outlets, but spread on online site. People can search the latest model of Islamic fashion online and choose the ones as they wish. However, the most frequent problem in searching the clothes online is the difficulty in describing the clothes model into keywords so that the searching result is not as expected. Therefore, Content Based Image Retrieval (CBIR) system is used to search the clothing based on the content of an image [1]. In CBIR, the image content is represented by feature, such as color, shape, and texture [2].

The interest of the people in fashionable clothes is not only in their color but also the model. As our concern is to obtain a particular model, the shape feature is used. The study of CBIR using shape feature has been done in various fields such as agriculture, medicine, advertisement, journalistic, and fashion. For example, in [3] the Curvature Scale Space (CSS) was used to extract the shape of the sea creatures. The CSS is the standard shape feature defined on Motion Picture Experts Group-7 (MPEG-7). The work showed that the CSS was simpler yet superior to Fourier based shape extraction method.

In the last decades, fashion items are becoming a popular research object in CBIR [4-7]. The work in [4] implemented the image RGB and SIFT with Bag of Key (BoK) histogram as the color features [4]. Before the method extracted the feature, the clothing image was divided into four areas, which are outside, inside, bottom, and shoes. Then, the RGB and BoK histograms were created based on the mentioned areas. The accuracy of this method was about 21.7%. Another CBIR study used the shape context to improve the clothes retrieval result [5]. The retrieval performance of the method was 32.73% with the average matching time was 2.4 minutes.

The next study applied the Histogram of Oriented Gradient method (HOG) on clothes classification [6]. In this study, the clothes image is grouped based on textile's pattern, the length of clothes or sleeve, the shape of the dress, etc. Each group of clothes contains different clothes pattern, such as animals, flowers, and lines. After feature extraction, the method was trained for each group and followed by testing. Furthermore, the HOG descriptor was applied to detect human [7]. It was reported that the performance of the HOG was better than wavelet.

The Pyramid Histogram of Oriented Gradient (PHOG) is an approach to extract shape feature from an image, was previously applied to CBIR of animal, plant, and fruit images [8]. In addition, PHOG method has also been implemented as the character feature extractor on an Optical Character Recognition (OCR). Various testings were conducted, and the accuracy level was 75% to 82% [9].

This article proposed a new normalization procedure to improve the performance of Islamic women clothing retrieval, which is conducted before applying the PHOG method. The contributions of the proposed method are as follow:

- a. Discuss the implementation of original PHOG method as an approach to extract the shape feature of Islamic women clothes. Based on an extensive study literature, the PHOG has not been applied to extract the shape feature of Islamic women's clothing. As in common, this clothing style has the different appearance to other styles, the discussion and implementation have become an important issue.
- b. Propose a new grid normalization before implementing the PHOG. We refer to our proposed method as normalized PHOG (NPHOG).
- c. Show that the performance of the NPHOG is better than PHOG and HOG, in the case of Islamic women clothes retrieval.

## 2. Pyramid Histogram of Oriented Gradients (PHOG)

Pyramid Histogram of Oriented Gradients (PHOG) is a method to extract shape feature of an image. PHOG is the representation of pyramid from HOG descriptor or the combination of HOG features [8]. There are three steps in the PHOG feature extraction process: (1) calculate the gradient, (2) decide the bin orientation to create the histogram, and (3) create the spatial pyramid.

### a. Gradients Calculation

The gradient calculation is done to have horizontal and vertical gradient direction from the object in an image. In this study, we use the following Sobel filter to calculate the gradient of an image.

$$K_X = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} K_Y = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

The horizontal gradient matrix ( $G_x$ ) is acquired from the convolution of  $K_X$  filter with the grayscale image, while the vertical gradient matrix ( $G_y$ ) is obtained from the convolution of  $K_Y$  filter with the grayscale image. The next step is calculating the gradient direction of an object (pixels), according to (1).

$$\theta = \tan^{-1} \frac{G_y}{G_x} \quad (1)$$

### b. Histogram of Bin Orientation

The gradients  $\theta$  of each pixel in an image are grouped into nine bins, to create a histogram of bin orientation. The bin groups based on orientation are:  $0^\circ$ - $20^\circ$ (bin 1),  $21^\circ$ - $40^\circ$ (bin 2),  $41^\circ$ - $60^\circ$ (bin 3),  $61^\circ$ - $80^\circ$ (bin 4),  $81^\circ$ - $100^\circ$ (bin 5),  $101^\circ$ - $120^\circ$  (bin 6),  $121^\circ$ - $140^\circ$ (bin 7),  $141^\circ$ - $160^\circ$ (bin 8) and  $161^\circ$ - $180^\circ$ (bin 9).

### c. Spatial Pyramid

Figure 1 shows the process of spatial pyramid creation. First, the image gradients are calculated, then the image is divided into three levels, which are level 0, 1, and 2. The level 0 consists of one grid, level 1 is divided into four grids, while level 2 is divided into 16 grids. The histogram is then generated from all grids at all levels. The PHOG feature is the concatenation of these individual histograms.

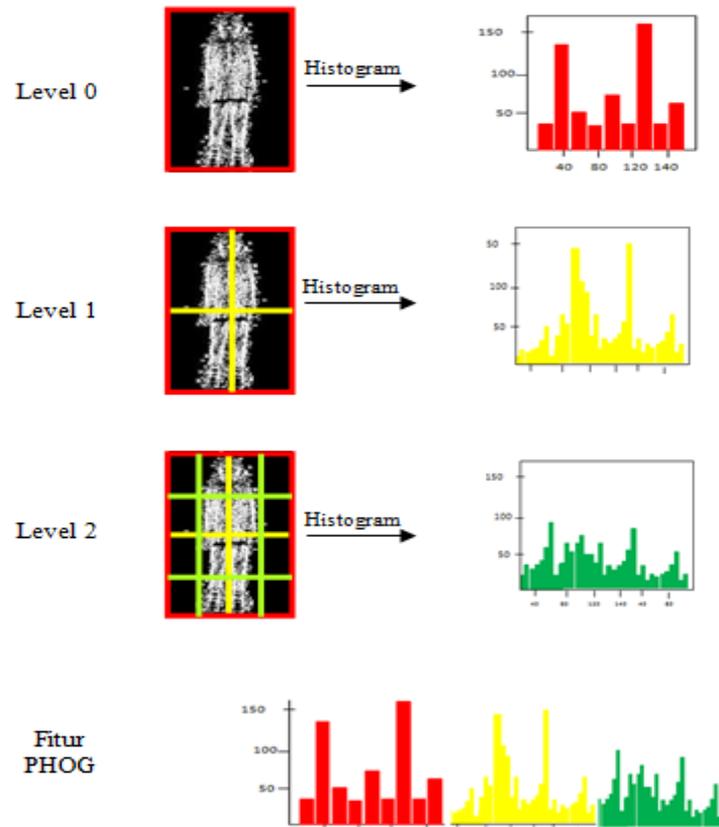


Figure 1. PHOG spatial pyramid

**3. Research Method**

The object used in this study are images of Islamic women apparels. These images consist of four groups, which are blouse-pants, long dress, outerwear, and tunic. Each group contains 85 images with the size of 136 x 212 pixels, and white background. This size is considered suitable so that the PHOG can be extracted quickly but at the same time sufficiently representative. The image was collected from the online catalog such as hijabenka.com and saqina.com. Some examples of the apparels can be seen in Figure 2.



Figure 2. Examples of Islamic women apparel

Next, we will explain the process of the CBIR in Islamic women apparels by using NPHOG as feature extraction. Figure 3 shows the CBIR system that consists of the query and data base sides and can be accomplished in three stages: pre-processing, NPHOG extraction and calculation of feature similarity.

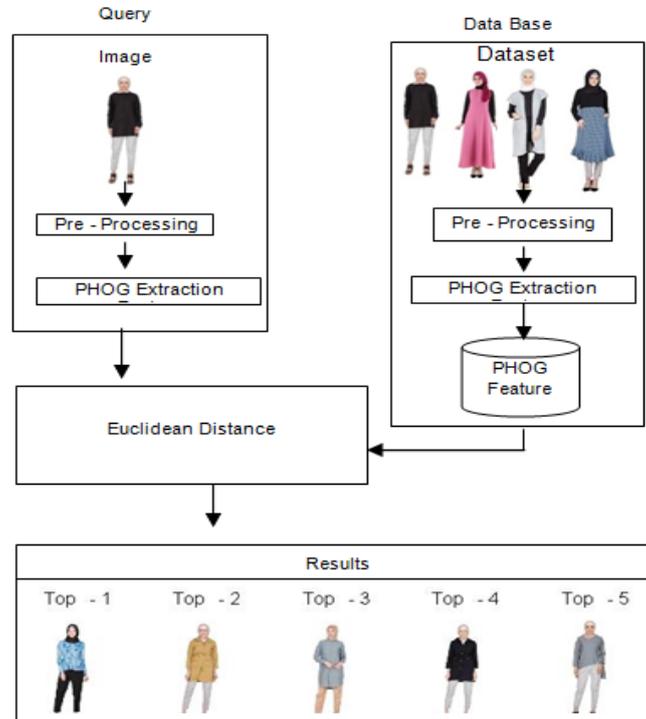


Figure 3. CBIR workflow on Muslimah clothes

#### a. Pre-processing

Pre-processing is the first stage of the CBIR. It is the process of image preparation to be able to generate the best feature. In this stage, the RGB image either from query or data base will be converted to a grayscale image. Changing the color image to grayscale format is done by averaging the pixel value of red, green and blue component [10], according to (2).

$$\text{Grayscale Image} = \frac{R+G+B}{3} \quad (2)$$

where R, G and B are the pixel values of the red, green and blue component, respectively. Then, we calculated the gradient and its direction of the grayscale image using (1).

#### b. NPHOG extraction

Figures 4 and 5 show the NPHOG spatial pyramid generation process. Figure 4 illustrates the first step that was detecting the face of an object using Viola Jones method; the center point of the face was set to be the reference for normalization.

Figure 5 illustrates the proposed normalization of the spatial pyramid (NPHOG). Level 0 was set to one grid, level 1 was set into four grids namely upper, left, right, and center grids. Then each grid at the level 1 was divided into four grids, so the total number of the grid is 16. The NPHOG feature is obtained by combining the gradient histogram from level 0, 1, and 2. Each grid has one histogram. So, the total number of the histogram from those three levels was 21. With nine bins in each histogram, the NPHOG feature consists of 189 vector dimensions. Figure 6 shows the vector formation of the NPHOG feature extracted from all the grids.

We proposed a normalization approach that was based on the position of the model's face in the image. The coordinate of the face is the reference point of the garments. Even if this step is similar to the conventional grid partitions described in Figure 1, the proposed normalization resulted in different partitions as shown in Figure 5. For example, at the level 0,

the red frame covering the model is at the image center relative to model's face. The pixels outside the frame were not considered in histogram generation. Next, our results show that this approach increased the performance of clothing retrieval.



Figure 4. Face detection

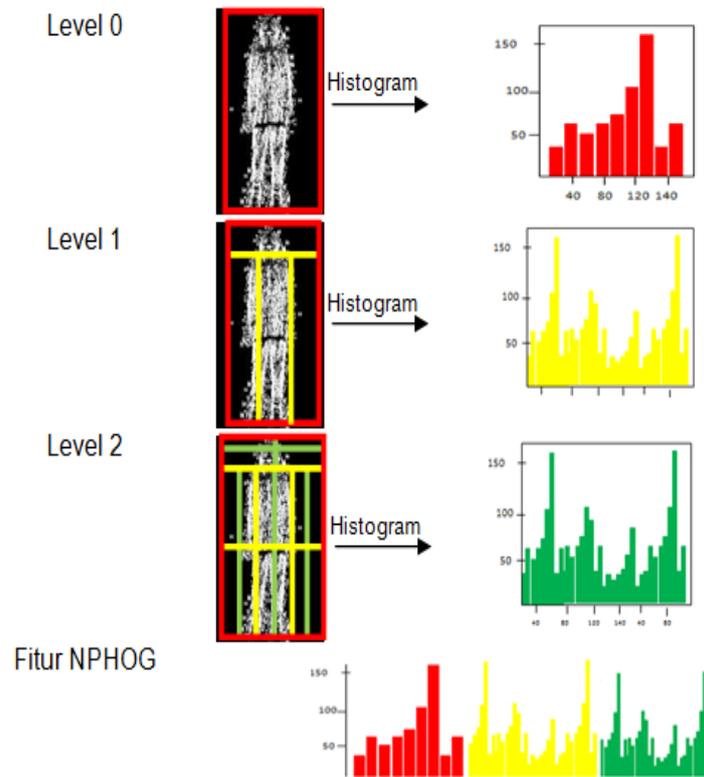


Figure 5. NPHOG Spatial Pyramid; the proposed normalization approach

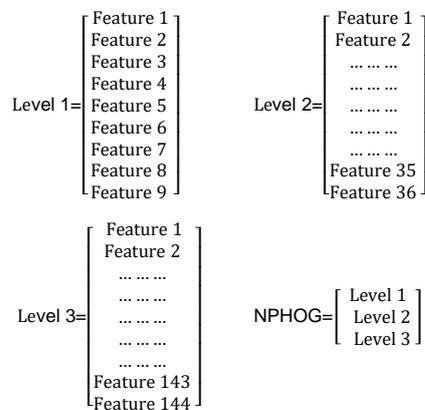


Figure 6. NPHOG vector dimension

c. The calculation of feature similarity

The next stage is calculating the similarity between query image with data set image feature using Euclidean distance method. It is a method that frequently used in calculating the similarity of two vectors. The Euclidean distance is given in (3).

$$d_{XY} = \sqrt{\sum_{i=1}^n (X_n - Y_n)^2} \quad (3)$$

where  $d_{XY}$  is the Euclidean Distance between vector  $X$  and  $Y$ ,  $n$  is the number of vector element,  $X_n$  is the  $n$ th element of vector  $X$  and  $Y_n$  is the  $n$ th element of vector  $Y$ .

The performance of the CBIR is measured using recall and precision. For each querying, the precision and recall are given by (4) and (5).

Precision is the comparison between the number of relevant images and the total number of retrieved image, which is given in (4) [11].

$$\text{Precision} = \frac{\text{The number of relevant image}}{\text{Total number of retrieved image}} \quad (4)$$

Recall is the comparison between the number of relevant image and the total relevant images in the data base, given in (5) [11].

$$\text{Recall} = \frac{\text{The number of relevant image}}{\text{Total number of relevant image in database}} \quad (5)$$

#### 4. Results and Analysis

In this section, we present the implementation results of PHOG extraction at the pre-processing step and NPHOG extraction step. Next, we explain the retrieval results of each dress group, followed by the discussion of impact of the proposed grid normalization. Performance of the NPHOG was compared to those of the PHOG and HOG and was evaluated by using 340 clothing images, comprised of four clothing categories, 85 images for each category: blouse-pants, long dress, outerwear, and tunic. We chose 5 query images randomly and accomplished five simulation cycles in each group.

##### 4.1. PHOG Extraction Process

Figure 7 shows the image conversion process from RGB to grayscale and visualization of horizontal and vertical images and gradient direction image. From the gradient image in Figure 7(d), the NPHOG feature was generated. The number of bin used to create the histogram was nine bins as described in section 2.b

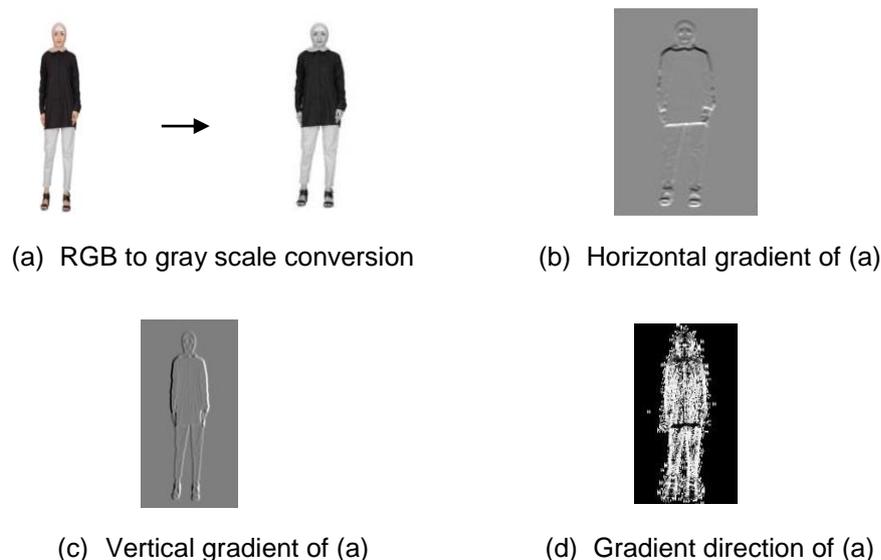


Figure 7. Visualization example of horizontal and vertical edge images, and gradient image

#### 4.2. Results of CBIR in Each Apparel Group

The image retrieval result examples on Islamic dresses using NPHOG are shown in Figure 8 and 9. Each of the figures shows 20 images that have the closest Euclidian Distance (ED) to the query image. The query image is shown in the top-left corner; while the 'top-1' is the retrieval result with the closest ED to the query image, which was the query image itself; 'top-2' is the 2nd closest to the query image, and so on. The retrieved image that was not relevant to query image is marked by the red square.

Figure 8 shows a result example of blouse-pants retrieval, an image from the tunic group was retrieved at the 14th position. From the five simulation cycles, it turned out that the most irrelevant retrieved image was from the tunic group.

Figure 9 shows a result example of the long dress retrieval, the images at the position of number 11, 15, 16, 18, and 20 were irrelevant images to the query image. Those images were tunic (at the 'top-11') and outerwear (at the 'top-15', 16, 18 and 20). An observation of the results indicated that the NPHOG was not able to differentiate the dress detail positioned in the center. For example, the blazers at the position 'top-15' and 'top-20', which have the different detail in the center of the dress, were assumed to be long-dresses.



Figure 8. Results of retrieved images in blouse-pants group



Figure 9. Results of retrieved images in long-dress group

**4.3. The Impact of NPHOG Normalization**

The line graphs in Figure 10 are the results of NPHOG-based retrieval for all dress groups. Eight points on each line show the intersection of recall and precision value for the different number of retrieval done in the simulation. The left to right points indicate the 10, 20, 30, 40, 50, 60, 70 and 85 retrieved images. For all retrieval number, the group of blouse-pants shows the best retrieval performance, followed by long dress, outerwear, and tunic.

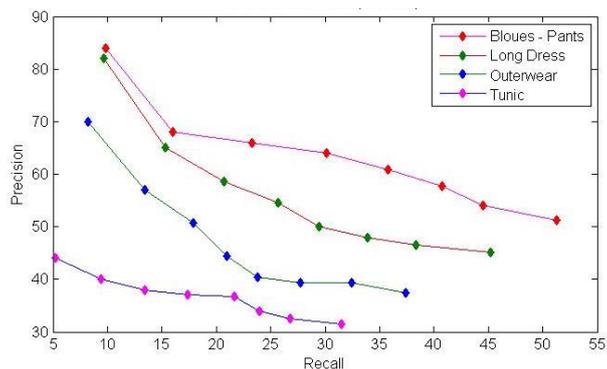


Figure 10. Recall and precision of NPHOG of all apparel groups

The shape (outline) of blouse-pants is different from that of the other groups, except tunic; on the other hands, long dress, outerwear, and tunic have a similar shape. Thus, if the query is from blouse-pants, the retrieved images will be likely from the same group. However, if the query is long-dress, outerwear or tunic, the images from other groups, except blouse-pants,

may be retrieved. In this case, similar outline of dresses in blouse-pants group yielded similar gradient direction ( $\theta$ ), thus low ED value, and finally higher recall and precision values.

Figure 11 shows the HOG, PHOG and NPHOG performances of four dress groups. With some exceptions, the NPHOG performed better than PHOG and HOG in three dress groups namely blouse-pants, outerwear, and tunic. However, for the long dress, the HOG and PHOG were better than NPHOG. In the first three groups, the apparel shapes were more complex than that in the long-dress group. Here, the grid normalization applied in the NPHOG resulted in more representative features, thus can achieve the higher recall and precision values.

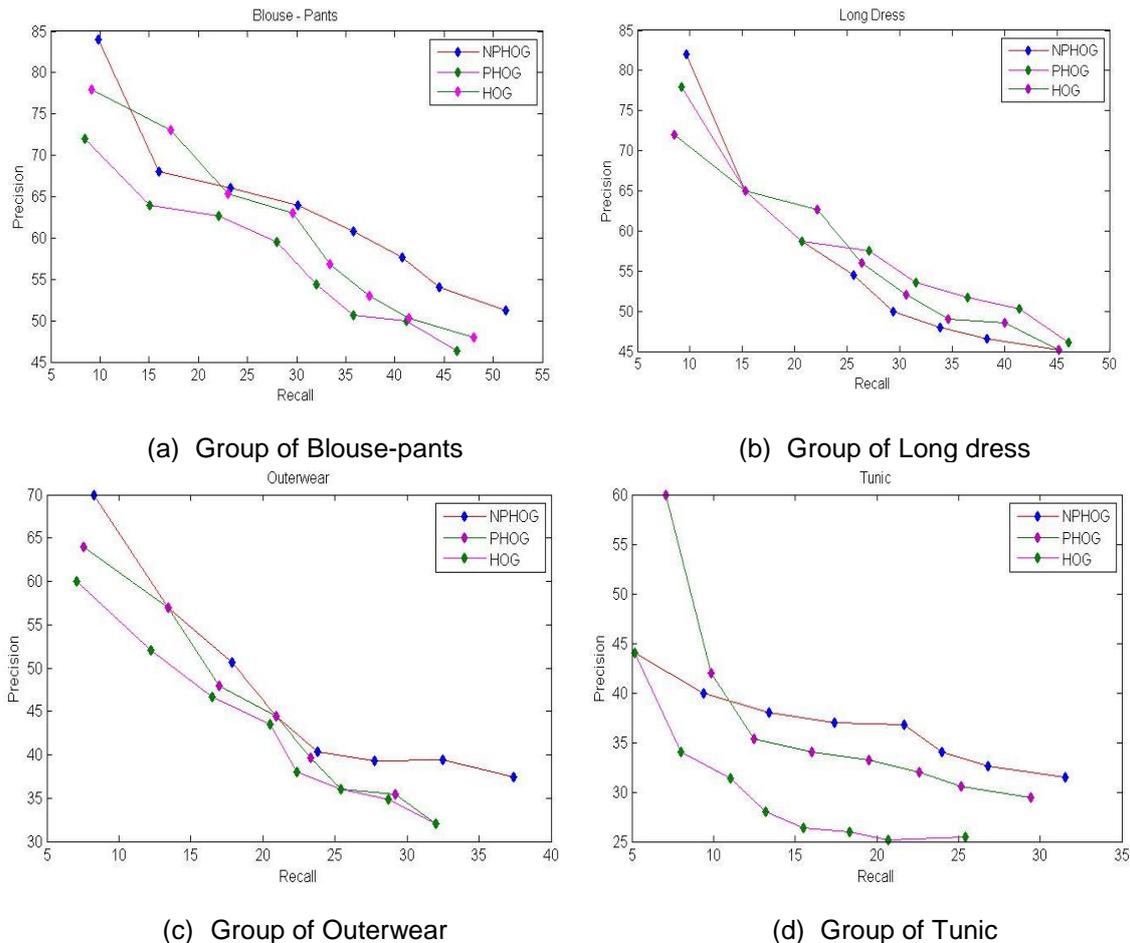


Figure 11. The recall and precision of NPHOG, PHOG and HOG in four dress groups

Averaging the precision values from Figure 11, it was obtained that for the group of blouse-pants, the precision value of NPHOG was  $\pm 6\%$  higher than that of PHOG and  $2\%$  higher than that of HOG method. In the outerwear group, the precision of the NPHOG was also  $\pm 3\%$  higher than PHOG and  $4\%$  higher than HOG. Lastly, the precision of the NPHOG in tunic group was  $\pm 7\%$  higher than HOG method.

**5. Conclusion**

The NPHOG method was successfully applied in extracting the shape feature of Muslim apparels in a dataset with 340 images grouped into four classes: blouse-pants, long dress, outerwear, and tunic. We proposed the grid normalization at level 1 and showed that it could improve the retrieval performance. The reference point of the grid normalization was the center point of the face that was detected using Viola-Jones method. Based on the recall and

precision, it showed that NPHOG was better than PHOG and HOG in three dresses group: blouse-pants, outerwear, and tunic. While in the long dress, the performance of NPHOG was lower than other features because the long dress shape was simpler, so the grid normalization did not increase the retrieval performance.

### Acknowledgments

This research is funded by the Ministry of Research, Technology and Higher Education of the Republic of Indonesia, under the Hibah Pascasarjana Scheme.

### References

- [1] N Singhai, SK Shandilya. A Survey on: Content Based Image Retrieval Systems. *International Journal of Computer Applications*. 2010; 4: 22-26.
- [2] PS Hiremath, Jagadeesh P. Content Based Image Retrieval Using Color, Testure and Shape Features. *International Journal of Computer Science and Security*. 2007; 1(4): 25-35.
- [3] Sadegh A, Farzin M, Josef K. Curvature Scale Space Image in Shape Similarity Retrieval. *Multimedia systems*. 1999; 7(6): 467-476.
- [4] Shinya M, Toshihiko Y, Kiyoharu A. *SNAPPER: Fashion Coordinate Image Retrieval System*. International Conference on. IEEE, 2013.
- [5] Chin HT, Shou SH, et al. An Efficient Garment Visual Search Based on Shape Context. *Journal WSEAS Transactions on Computers*. 2009; 8(7):1195-1204.
- [6] Andrada GR. Attribute Based Image Classification and Retrieval. Thesis. Swiss: Swiss Federal Institute of Technology Zurich; 2014.
- [7] Nafneet D, Bill T. *Histograms of Oriented Gradients for Human Detection*. IEEE Computer Society Conference on Vision and Pattern Recognition, San Diego. 2005; 1:886-893.
- [8] Anna B, Andrew Z, Xavier M. *Representing Shape with a Spatial Pyramid Kernel*. Proceedings of the ACM International Conference on Image and Video Retrieval. Amsterdams. 2007: 401-408.
- [9] Zhi RT, Shangxuan T, Chew LT. *Using Pyramid of Histogram of Oriented Gradients on Natural Scene Text Recognition*. IEEE International Conference on Image and Video Retrieval. Paris. 2014: 2629-2633.
- [10] Candra NS. Mengubah Citra Berwarna menjadi Citra Grayscale dan Citra Biner. *Jurnal Teknologi Informasi DINAMIK*. 2011; 16(1): 14-19.
- [11] Shirahatti, Nikhil V, Kobus Barnard. Evaluating Image Retrieval. Thesis. Arizona: University of Arizona; 2005.