2019 IEEE International Conference on Cybernetics and Computational Intelligence (IEEE CYBERNETICSCOM),
Banda Aceh - Indonesia, August 22-24, 2019

HOG of Region of Interest for Improving Clothing Retrieval Performance

Niza Aulia, Fitri Arnia* and Khairul Munadi
Department of Electrical and Computer Engineering, Universitas Syiah Kuala
Email: *f.arnia@unsyiah.ac.id

***Abstract*—Due to spread of garment online store and the revenue from the online clothing market, clothing retrieval is becoming an active research topic in the last decade. The buyer commonly searches for shape, but the availability of various clothing designs and details on the Internet caused a growing interest to buy clothing with similar detail, at a particular position. The proposed method aims at enhancing the performance of garment image retrieval, in which the garments have similar details at a particular position. The method is developed based on histogram of oriented gradient (HOG), with a new approach for ROI partition mechanism prior employing the HOG. Simulations with four categories of clothing details confirmed that the proposed approach can improve the retrieval performance.**

***Index Terms*—clothing image similarity, clothing details, histogram of gradients, body proportion**

I. Introduction

Due to spread of garment online store and the revenue from the online clothing market, clothing retrieval is becoming an active research topic in the last decade. An online store should have an attractive representation, and handy tools for items searching should be provided to enhance the users shopping experience. Particularly for clothing, the clothing image feature used for retrieval should represent the users desires, such as shapes and details. Examples of garment shape are dresses, pants, skirts, jackets, and tunic; and examples of detail are pockets, pleat, waist bow, collar, etc. The buyer commonly searches for shape, but the availability of various clothing designs and details on the Internet caused a growing interest to buy clothing with similar detail, at a particular position.

Recently, several studies have been discussing the clothing retrieval system that concern about some details in the garments, such as collar and skirt shapes and details [1]–[3]. For retrieving the garment details, we need to design a particular feature capable of capturing the details; and this task is not trivial. Numerous attempts have been made to develop suitable features for clothing retrieval, and the histogram of oriented gradient (HOG) and pyramid HOG (PHOG) are some of those features [4]–[6]. The HOG and PHOG can achieve good retrieval results as long as the region of interest (ROI) is reasonably segmented. However, this is not the case, the PHOG is commonly applied traditionally, and this approach may cause some drawbacks in clothing retrieval. Figure 1 illustrates edge image partition in PHOG extraction process. First, the edge image is produced, which is the zero-level edge image pyramid; marked by the cyan box. From this image, the histogram of oriented gradient is built. The image is then partitioned into four equivalent sizes (area shown by red boxes) produced the first-level image pyramid. Subsequently, the histogram is generated from each part. Then, each red-boxed area is divided into four equivalent regions (shown by green-boxes) resulted in the second level image pyramid, and the histogram is also built from each part. Concatenation of these histograms is called the PHOG.

For representing the common images such as natural image [7], this approach has shown a moderate success. However, the conventional partition may mislead overall similarity of two garments. For illustration, in Fig. 1 grid 13, there is an approximate vertical edge. When comparing only this grid with its corresponding grid in another image, whose edges is relevant, but the clothing object is not —for example edge of a skirt— may result in a high similarity. For this reason, we suppose that a Region of Interest (ROI) should be appropriately defined and partitioned to maximize the image representation saved in a feature.

We proposed a garment retrieval method based on a new concept of ROI. Based on a rigorous observation of woman's wardrobe, we concluded that generally the details are located in three parts of the clothes: upper parts such as collar and tie, middle part such as belt with various form, and lower part including pleat with its variations. We proposed a mechanism to partition the garment into the upper, middle, and lower parts by using the rule of seven-head height originally used in human body anatomy and art. The upper, middle and lower parts are denoted as ROI 1, ROI 2 and ROI 3 respectively. The HOGs were calculated from ROI 1, ROI 2, and ROI 3, presenting the image features. We conducted simulations with 200 woman's clothings.

II. Related Works

Retrieval of garment details has attracted some scholars, for examples the authors in [1]–[3] proposed the retrieval based on collar and skirt shapes and details. Shimizu et al. proposed a keen algorithm to retrieve collar shape and details [2]. Mizuochi proposed a system to retrieve desired garments that are globally similar to an image and partially similar to another image [1]. Kondo et al. proposed a method to retrieve skirts with similar shape and a particular detail at a specific location; such as a cloth bow at the right upper side of the skirt [3]. The method detected one-point decoration, i.e. the garment detail that is detected by using the saliency map. In addition, they

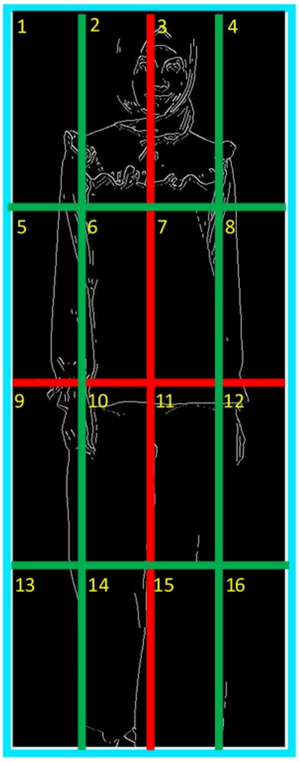

Fig. 1. Edge image partition in PHOG.

applied Fourier descriptor and Gabor filter to extract the shape and texture features respectively.

Concept of histogram of oriented gradient (HOG) was first introduced by McConnell [8]. However, it had not received much popularity before Dalal [9] applied the HOG to detect human in video. To overcome limitation of the HOG, a pyramid HOG was introduced in [7]. Thus far, the pyramid histogram of oriented gradient (PHOG) is one of shape features used frequently in numerous applications, such as emotion recognition [10], smile recognition [11], machine-printed/handwritten latin and Arabic word recognition [12], traffic sign detection [13], natural scene text recognition [14], human action recognition [15], plant diseased leaf segmentation and recognition [16], and clothing retrieval purpose [5].

For clothing description and retrieval, the HOG and PHOG has also been widely used [4]–[6]. Sha used PHOG, Fourier, and GIST features representing sleeve and collar. Cushen employed the HOG as part of the histogram of visual words to extract garment shape. Both Cushen and Sha considered only upper part of garments. Navarro used HOG as the descriptor of collar and necktie, which have been detected previously and marked as ROI. Navarro showed that the HOG performed better than Linear Binary Pattern (LBP). Navarro used armpit and chin location to locate the ROI. Finding these two items in an image is not a trivial problem due to variation in illumination, pose etc.

III. Proposed Method

The proposed method aims at enhancing the performance of garment image retrieval based on HOG. Here we proposed a new ROI partition mechanism prior employing the HOG. Given the query, the focus was to retrieve the garments with similar details at particular positions. Figure 2 shows the retrieval pipeline in which the proposed approach is represented in yellow boxes. The proposed approach comprises of two parts: query and database. At the query side, the input image is passed to two stages of pre-processing procedures: image conversion into gray-scale and normalization. From the normalized image, we identified the ROI using our proposed scheme. Subsequently, the edge was detected by using the Canny algorithm, and based on information from the ROI identification step, ROI was mapped to the edge image. Finally, the HOG of nine bin was generated from each ROI; forming the feature of query image.

The same procedures (boxed in the dashed line) were accomplished in the database for each garment image to obtain the database features. Then, the Euclidean distance (ED) of query and database features were calculated for all images in the database and ranked. The ED is given in 1. The retrieved images were the images in the database whose features had the closest distance to that of the queries.

$$d_{ij} = \sqrt{\Sigma_{k=1}^{n}(x_{ik} - x_{jk})^2} \tag{1}$$

where x_i and x_j are the feature vector of the query and the database image respectively, and d_{ij} is the distance between the vectors. The two images with the closest distance are considered the most similar, and those with the largest distance are considered the least similar.

A. Region of Interest (ROI) Identification

We proposed a novel concept in defining the ROI, which is illustrated in Fig. 3. First, the garment position was normalized based on the face location. This approach is similar with the technique described in [17]. The face location was detected by using Viola-Jones algorithm. Based on the face location, an approximate garment width was determined, shown by the blue dashed line. Subsequently, we applied the rule known as the seven-head height body proportion to partition the garment length into seven equal segments. We added one-fourth of the face height to cover the overall size of the head. Seven boxes of the head size are shown by the red boxes, each is numbered by ‘1’, until ‘7’. Based on the height of each box, the ROI segments were defined. The upper part of the garments is defined by the position and height of the box number 2; that is ROI 1. The middle part of the garment is determined based on position and height of the boxes number 3 and 4; that is ROI 2. And the ROI 3 is the rest of the boxes, which locates the garment lower part.

B. Histogram of Oriented Gradient of ROI

Traditionally, the pyramid histogram of oriented gradient (PHOG) is obtained from three (or more) pyramid level

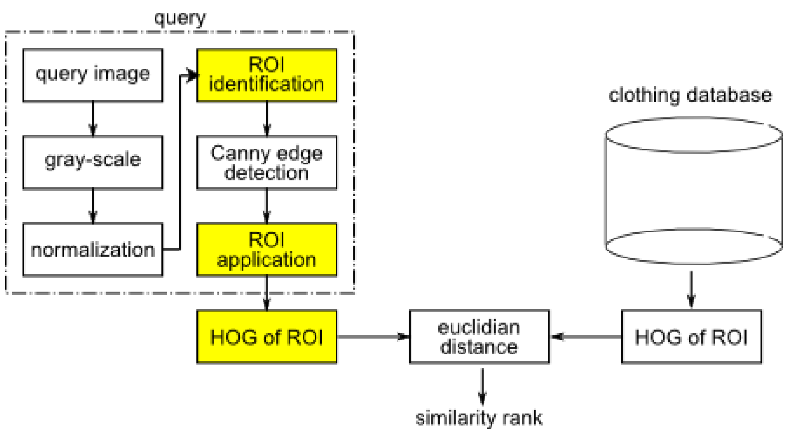

Fig. 2. The retrieval pipeline; the yellow boxes represent the proposed approach.

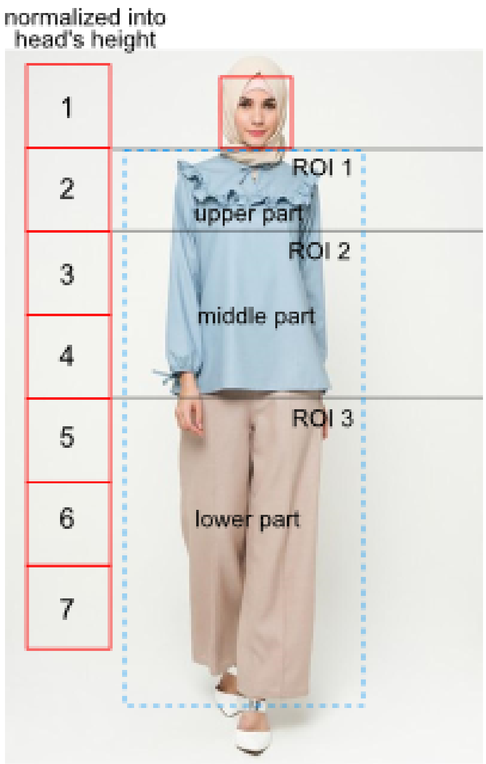

Fig. 3. The proposed ROI identification.

explained in Fig. 1; or HOG can be also build based on only the second level pyramid, i.e. the 16 uniform parts of the edge image. The proposed method built the histogram of oriented gradient from each ROI defined in Sect. III-A. The traditional PHOG or HOG has shown promising performance in many applications as discuss in Sect. II. However, the simulations showed that for clothing image retrieval, particularly those with details, the HOG of the proposed ROI enhance the retrieval performance of PHOG.

IV. Simulation and Results

A. Simulation Condition

The simulation was conducted using 200 women's clothing images as the object, that were obtained from several online shop websites. All image was resized into 800x1200 pixels. We chose the size to guarantee that all the details, which correspond to the edge, can be captured appropriately. There were four image categories based on the ROI: garments with details in upper, middle, and lower parts, and those with detail in more than one location, such as details in upper and middle part. We selected five queries from each category, thus, there were 20 query images. As we focused on the details and their respective location in the garments, we didn't consider the global shape. Thus, we classified the dresses and the pants with similar belt variations as one category; blouses and dresses with similar collars were also classified as one category. Some examples of clothing images used in the simulations are shown in Fig. 4.

B. Accuracy Measurement

We calculated average Precision values from five query images. Precision is defined in 2.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

here, TP and FP stand for true positive and false positive respectively.

C. Results and Discussion

Figures 5 and 6 show the clothing retrieval results with details in the upper and middle parts of the garments respectively; (a) results of the proposed method, and (b) PHOG. Here, the PHOG described in Fig. 1 was applied to the clothing images excluding the model's head. The results of PHOG was used as a based for measuring the proposed method performance. Both figures display ten database images whose feature distances were the closest to the queries, and the red boxes frame the clothing images correctly retrieved (true positives). The most similar clothing is marked with "Top-1", the second similar clothing is marked with "Top-2" and so on. The "Top-1" is

Fig. 4. Examples of several clothing images used in the simulation.

Fig. 5. An example of clothing retrieval results of the proposed method and PHOG with details in the upper part.

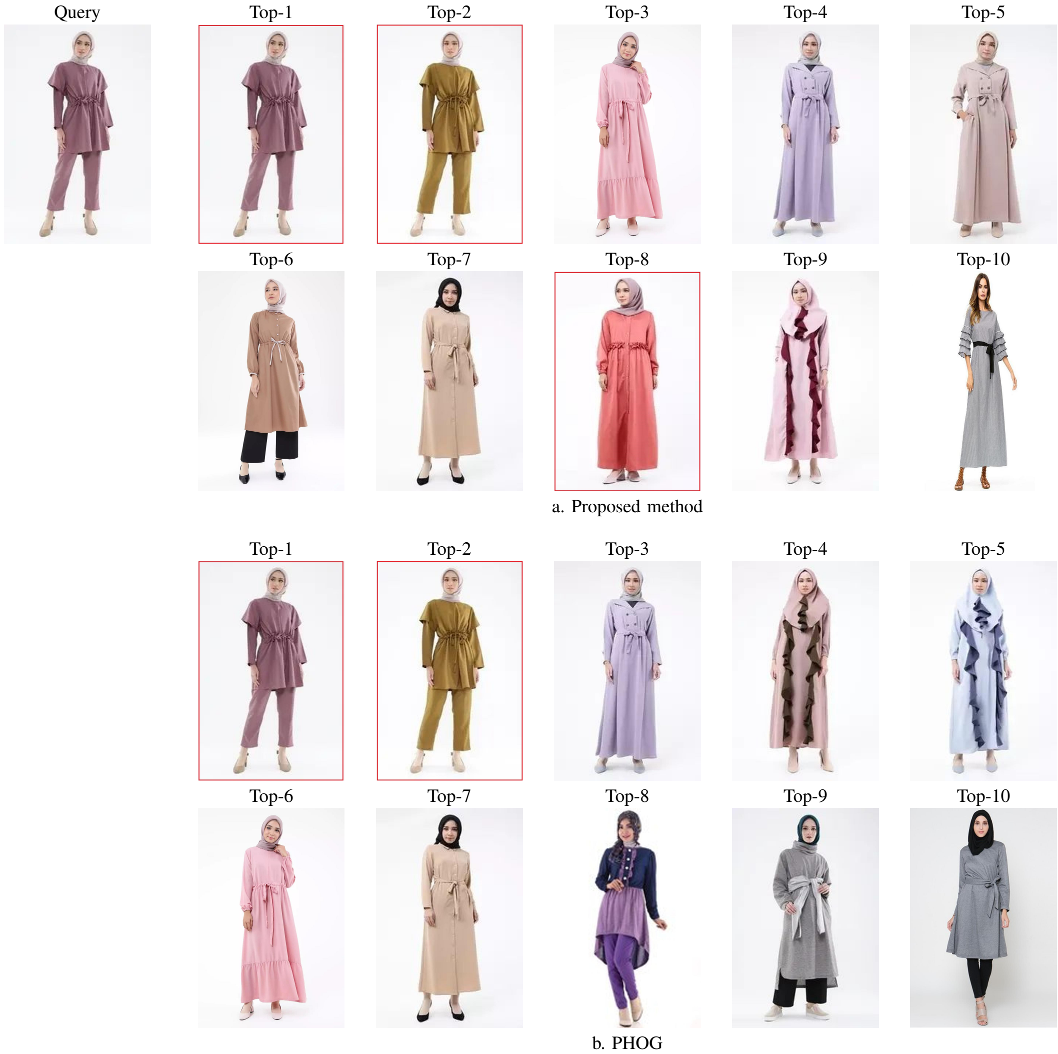

Fig. 6. An example of clothing retrieval results of proposed method and PHOG with details in the middle part.

the image with zero distance to the query, that was the query image itself.

From the figures, we can observe that the proposed method retrieved more correct images; it achieved higher precision values than PHOG. In Fig. 5, the proposed method retrieved four clothing images correctly, while the PHOG retrieved two images. In Fig. 6 the proposed method retrieved three clothing images, while the PHOG retrieved only two images respectively. In simulation, we applied a strict similarity rule on precision calculation; the garment with waist ruffles was not assigned the same class as waist ruffle with a bow. In case that we apply a looser rule, i.e., waist ruffles was assigned the same class as waist ruffle with a bow; the proposed method retrieved more relevant images, with only one false positive ("Top-9" in Fig. 6a). While the PHOG retrieved more false positives: "Top-4", "Top-5", and "Top-8" (see Fig. 6b) .

We showed some results in which the proposed method retrieved more relevant images as compared to the PHOG. Fortunately, these were the case for most of the query images. On average, the precision values of the proposed plan were higher than those of the PHOG. Figure 7 shows the average precision values of five queries from each clothing category. In the figure, the dashed lines represent the precision of the proposed method, while the solid lines represent those of PHOG. On each line, there are points referring to the number of retrieved image. For brevity, the point indicates n retrieved images is denoted by R-n; so the point of ten retrieved image is denoted by R-10 and so on.

Figure 7 shows that the proposed method moderately improved the precision values in all categories. Here, the trends of the proposed method were in line with, but higher than those of the PHOG; all dashed lines were higher than solid lines, with some exceptions. The precision values were increased between 1.5% to 16%; 1.5% increment was reached at R-80, for garments with upper details, while the 16% increment was achieved at R-10, for garments with details in lower parts. The proposed method showed a fair precision improvements for garments with details in more than one location (gray lines).

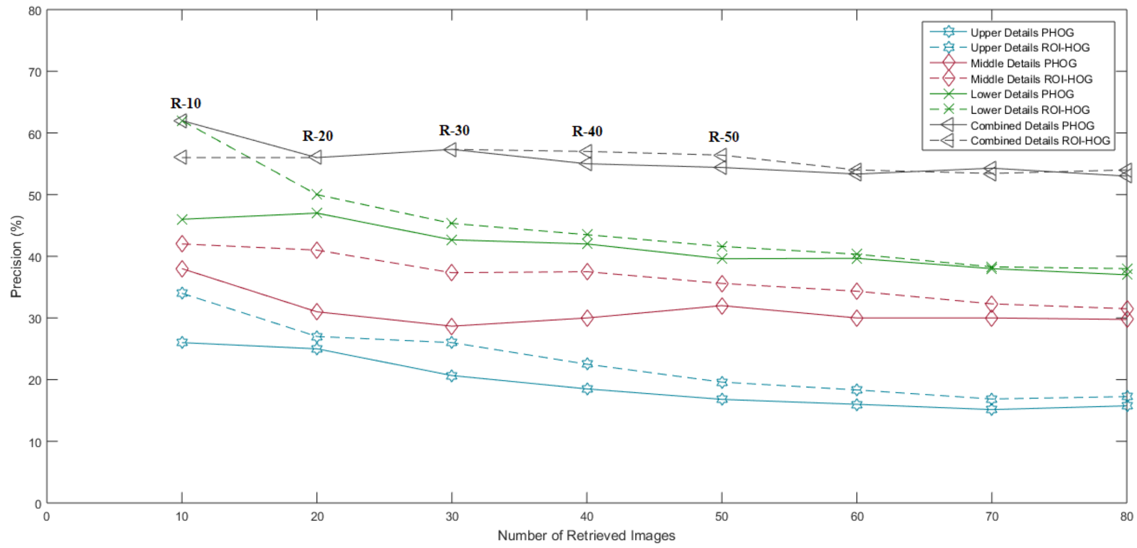

Fig. 7. Average precision values of five queries from each clothing category.

Among four categories, the proposed method achieved a more steady, and higher precision values for garments with detail in the middle part (see red lines). It is worth noting that the precision values of the proposed method, and those of the PHOG were the lowest for the images with upper details. One factor that may cause this was the data set consisted of the woman with headscarf, which presumably appeared to be the upper details.

V. Conclusion

We proposed a method aims at enhancing the performance of garment image retrieval, in which the garments have similar details in particular positions. The proposed method was developed based on HOG, with a particular concern to define a new ROI partition scheme. We proposed a mechanism to partition the garment into the upper, middle, and lower parts by using the rule of seven-head height originally used in human body anatomy and art. Simulations with four categories of clothing details confirmed that the proposed approach can improve the retrieval performance.

Acknowledgment

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

References

[1] M. Mizuochi, A. Kanezaki, and T. Harada, "Clothing retrieval based on local similarity with multiple images," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 1165–1168.

[2] K. Shimizu, W. Yang, M. Toyoura, and X. Mao, "Relevance feedback based retrieval of cloth image with focus on collar design," in *2015 International Conference on Cyberworlds (CW)*. IEEE, 2015, pp. 137–144.

[3] S.-I. Kondo, M. Toyoura, and X. Mao, "Sketch based skirt image retrieval," in *Proceedings of the 4th Joint Symposium on Computational Aesthetics, Non-Photorealistic Animation and Rendering, and Sketch-Based Interfaces and Modeling*. ACM, 2014, pp. 11–16.

[4] J. Lorenzo-Navarro, M. Castrillón, E. Ramón, and D. Freire, "Evaluation of lbp and hog descriptors for clothing attribute description," in *International Workshop on Video Analytics for Audience Measurement in Retail and Digital Signage*. Springer, 2014, pp. 53–65.

[5] G. A. Cushen and M. S. Nixon, "Mobile visual clothing search," in *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. IEEE, 2013, pp. 1–6.

[6] D. Sha, D. Wang, X. Zhou, S. Feng, Y. Zhang, and G. Yu, "An approach for clothing recommendation based on multiple image attributes," in *International Conference on Web-Age Information Management*. Springer, 2016, pp. 272–285.

[7] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *Proceedings of the 6th ACM international conference on Image and video retrieval*. ACM, 2007, pp. 401–408.

[8] R. K. McConnell, "Method of and apparatus for pattern recognition," Jan. 28 1986, uS Patent 4,567,610.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *international Conference on computer vision & Pattern Recognition (CVPR'05)*, vol. 1. IEEE Computer Society, 2005, pp. 886–893.

[10] A. Dhall, A. Asthana, R. Goecke, and T. Gedeon, "Emotion recognition using phog and lpq features," in *Face and Gesture 2011*. IEEE, 2011, pp. 878–883.

[11] Y. Bai, L. Guo, L. Jin, and Q. Huang, "A novel feature extraction method using pyramid histogram of orientation gradients for smile recognition," in *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2009, pp. 3305–3308.

[12] A. Saïdani and A. K. Echi, "Pyramid histogram of oriented gradient for machine-printed/handwritten and arabic/latin word discrimination," in *2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*. IEEE, 2014, pp. 267–272.

[13] A. Sugiharto and A. Harjoko, "Traffic sign detection based on hog and phog using binary svm and k-nn," in *2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*. IEEE, 2016, pp. 317–321.

[14] Z. R. Tan, S. Tian, and C. L. Tan, "Using pyramid of histogram of oriented gradients on natural scene text recognition," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 2629–2633.

[15] J. Wang, P. Liu, M. F. She, A. Kouzani, and S. Nahavandi, "Human action recognition based on pyramid histogram of oriented gradients," in *2011 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2011, pp. 2449–2454.

[16] S. Zhang, H. Wang, W. Huang, and Z. You, "Plant diseased leaf segmentation and recognition by fusion of superpixel, k-means and phog," *Optik*, vol. 157, pp. 866–872, 2018.

[17] C. Mutia, F. Arnia, and R. Muharar, "Improving the performance of cbir on islamic women apparels using normalized phog," *Bulletin of Electrical Engineering and Informatics*, vol. 6, no. 3, pp. 271–280, 2017.