Online product search using gray level co-occurrence matrix, color moments, and histogram of oriented gradients for content based image retrieval

To cite this article: Henry Halim *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **852** 012140

Online product search using gray level co-occurrence matrix, color moments, and histogram of oriented gradients for content based image retrieval

Henry Halim*, J Hendryli, D E Herwindiati
Informatics Department, Faculty of Information Technology,
Universitas Tarumanagara, Jakarta 11440, Indonesia

*henryhalim97@gmail.com

Abstract. Usually an information retrieval system uses text as its query for retrieving the results. However, the system can also use images, as opposed to text, for the search query. This technique commonly is called content based image retrieval. In this paper, we explore the retrieval of fashion products such as hats, sneakers, t-shirts, flat shoes, and dresses from images using color, texture, and shape features to represent the characteristics of the query images. The color, texture, and shape features are color moments, gray level co-occurrence matrix, and histogram of oriented gradients, respectively. The feature extraction process produces vectors with statistical feature value which can then be clustered to retrieve images with high degree of similarity. Comparing those feature representations for the retrieval system, it is shown that shape feature is the best representation.

1. Introduction

Nowadays, the searching and retrieval system is no longer limited to textual data, since another form of data, such as image, can also be used in a retrieval system. The technique is commonly called content based image retrieval or CBIR for short [1]. Information searching based on visual data is deemed as more practical and user-friendly where there is no need to memorize or type the correct keyword. We aim to develop a mobile-based CBIR system for fashion product searching using clustering method. The feature extraction process takes out statistical features from an image using above methods and uses it for the clustering and retrieval process [2].

Three representation of the image is explored here, which is the color moments, gray level co-occurrence matrix (texture representation), and the histogram of oriented gradients (shape). For the color representation, the HSV image is used for input and the color moment vector of the image is then computed [3]. Furthermore, the texture representation of the image is extracted using gray level co-occurrence matrix method (GLCM) [4]. And finally, for the shape representation, the histogram of oriented gradients (HOG) of the image is extracted [5]. Both the GLCM and HOG method use grayscale images as the input.

The CBIR system developed in this research retrieves images of fashion products, such as hats, sneakers, t-shirts, flat shoes, and dresses. The products are collected from several online markeplace in Indonesia and cover male and female fashion products. There are 151 images for sneakers, t-shirt, and hats and 150 images for both flat shoes and dresses. The example of images in the dataset are shown in Figure 1.

IOP Conf. Series: Materials Science and Engineering **852** (2020) 012140 doi:10.1088/1757-899X/852/1/012140

Figure 1. Example of fashion product

2. CBIR feature extraction process

Feature extraction process is the first step in the CBIR system. In this process, the system extracts all image features into feature vectors. The features consist of color, texture, and shape representation of the images.

2.1. Gray Level Co-occurrence Matrix

Texture feature has a role to separate regions in the images. It refers to the visual pattern that have property of homogeneity that doesn't produce single color or intensity [6]. The texture representation is extracted using gray level co-occurrence matrix (GLCM). It is used to calculate the spacial dependence of gray levels in an image. The size of rows and columns of the GLCM matrix equals to the number of gray levels in the image. The GLCM matrix is constructed in four spacial orientations: 0^0, 45^0, 90^0, and 135^0.

Let GLCM matrix be $P_{i,j}$ and the size of the matrix is equal to the number of gray levels in the image. Each element of $(i,\ j)$ represents the frequency by which pixel has gray level i related to neighbouring pixels with gray level j. Construction of GLCM from a grayscale image is illustrated in Figure 2.

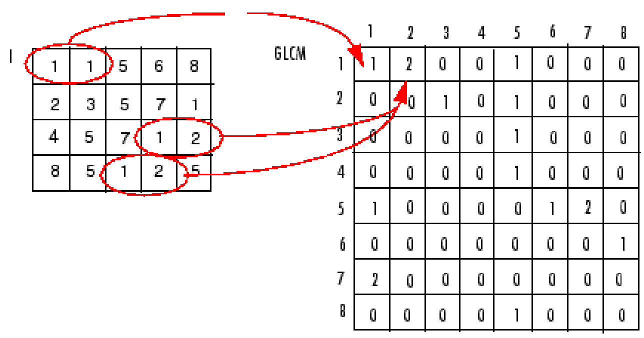

Figure 2. The illustration of the construction of GLCM matrix in 0^0 orientation [6]

The texture features from GLCM being used in this research are contrast, homogeneity, dissimilarity, energy, and angular second moment (ASM). The contrast measurement is usually called the local intensity variation and it can be calculated using Eq. 1[4].

$$\text{Contrast} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} |i-j|^2 \, P_{i,j} \tag{1}$$

Meanwhile, the homogeneity achieves its largest value when most of the occurrences in GLCM are concentrated near the main diagonal (Eq. 2)[4].

$$\text{Homogeneity} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1+(i-j)^2} P_{i,j} \tag{2}$$

The dissimilarity is the variation of gray level voxel pairs, and can be calculated as in Eq. 3[4].

$$\text{Dissimilarity} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P_{i,j} \, |i-j| \tag{3}$$

ASM feature is a measure of textural uniformity of an image. The energy reaches its highest value when grey level distribution has either a constant or a periodic form. The formula for calculating the ASM can be seen in Eq. 4[4].

$$\text{ASM} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{P_{i,j}\}^2 \tag{4}$$

And finally, energy is often called the square root of the ASM, which can be calculated as in Eq. 5[4].

$$\text{Energy} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{P_{i,j}\}^2 \tag{5}$$

2.2. Color Moments

Color moments are measures that can be used to identify images based on colors. Once calculated, these moments provide a measurement for color similarity between images. It can be used to compare the value between images collection in the database to the query image. In the color moments, there are three values that can be calculated: mean, standard deviation, and skewness. It should be noted that the color moment uses images in HSV colorspace as its input [3].

The mean is also known as the average color value in the image. It can be calculated as in Eq. 6[3].

$$\mathbf{Mean} = \sum_{j=1}^{N} \frac{1}{N} P_{i,j} \tag{6}$$

The standard deviation is the square root of the variance of the distribution (Eq. 7) where $\boldsymbol{E_i}$ is the mean[3].

$$\textbf{Standard Deviation} = \sqrt{\left(\frac{1}{N}\sum_{j=1}^{N}(P_{i,j} - E_i)^2\right)} \tag{7}$$

Lastly, skewness can be understood as a measure of the degree of asymmetry in the distribution (Eq. 8)[3].

$$\textbf{Skewness} = \sqrt[3]{\left(\frac{1}{N}\sum_{j=1}^{N}(P_{i,j} - E_i)^3\right)} \tag{8}$$

2.3. Histogram of Oriented Gradients
Histogram of oriented gradients or HOG is a very suitable descriptor for image-based retrieval system. Moreover, HOG descriptors are also widely used in computer vision, where the window-based descriptors used for object detection in an image. The HOG descriptor has a few benefits over other descriptor methods such as invariance to numerical and photometric [5]. The idea of this descriptor is divide the image into small related regions and calculate the edge intensity and gradients directions of the objects in these regions. This region is known as window which divided into $n \times n$ grid. Then, the descriptor computes the quantized q-bins histogram of edge intensity and gradient magnitude of each cells within a grid as in Eq. 9[5]:

$$G = \sqrt{Gx^2 + Gy^2} \tag{9}$$

where Gx and Gy denote the gradient of the image in x and y direction, respectively. For the orientation, it can be extracted as in Eq. 10[5].

$$\theta = \arctan\left(\frac{Gy}{Gx}\right) \tag{10}$$

The last step is to do a histogram calculation of the cell orientation counted before. Each pixels in a cell has its own histogram value based on the value generated in the gradient calculation.

3. Clustering & Retrieval Process
After the feature extraction process, the system clusters the feature vectors from the query image and image collections stored in the database. The k-means clustering method is employed in this step. K value used in k-means clustering method is the optimal value that obtained from the silhouette scoring experiment, we can use the silhouette scoring method to looking for the best k value for k-means clustering[7]. The system retrieves top 40 images of the same cluster in the database with the shortest euclidean distance to the query image. Table 1 shows the example of the image retrieval result for a fashion product search using only the shape feature of dress image.

4. Analysis & Evaluation
The experiment consists of four scenarios where we test the best combination of feature extraction methods for the retrieval process. The shape feature is employed in all scenarios, since it is logically the best representation of the fashion products. On each scenarios, we include color, texture, both color and texture, and also only shape feature. The evaluation uses mean average precision (MAP) metric which is calculated from the precision and recall of the retrieval result, as in Eq. 11, where $\boldsymbol{Q}$ is the number of queries, $\boldsymbol{N}$ is the number of retrieved images, $\boldsymbol{k}$ is the rank in the sequence of retrieved images, $\boldsymbol{P_q(k)}$

is the precision at cut-off $\boldsymbol{k}$ for query $\boldsymbol{q}$, and $\boldsymbol{\Delta r_q(k)}$ is the change in recall from items $\boldsymbol{k-1}$ to $\boldsymbol{k}$ for query $\boldsymbol{q}$. The result of all experiment scenarios is shown in Table 2.

$$\mathrm{MAP} = \frac{\sum_{q=1}^{Q}\sum_{k=1}^{N} \boldsymbol{P_q(k)\Delta r_q(k)}}{\boldsymbol{Q}} \tag{11}$$

The result of the experiment shown in MAP value is relatively low due to the recall value calculated for top 40 result. According to the result, we can get a conclusion that using the shape feature results in the best retrieval performance compared to other feature combinations. The shape of the product in the image is arguably enough to differentiate among the fashion products. But, it can also be attributed to the distribution and consistency of textures and colors of the products in the database. As we know, the textures and colors of the fashion products can be varied, so a large number of images may be needed for better retrieval result.

Table 1. Example of image retrieval result for a dress image

Query	Image Retrieval
	

Table 2. The result of four experiment scenarios

Scenario	**MAP Value**
Only shape feature	0.1303
Color & shape feature	0.0718
Texture & shape feature	0.0847
Color, texture, and shape feature	0.0957

References

[1] Layona R, Tunardi Y, Tanoto Image Retrieval Berdasarkan Fitur Warna, Bentuk, dan Tekstur *ComTech: Computer, Mathematics and Engineering Applications* **5** 1073-1085
[2] Karthikeyan T, P Manikandaprabhu, S. Nithya A survey on text and content based image retrieval system of image mining *International Journal of Engineering Research* **3** 509-12

[3] Keen N Color Moments *School of Informatics, University of Edinburgh* 3-6
[4] Shijin K.P.S, Dharun V.S Extraction of texture features using GLCM and shape features using connected regions *International Journal of Engineering & Technology* **8** 2926-2930
[5] Randa A.F, Suciati N, Navastara D.A Implementasi Metode Kombinasi Histogram of Oriented Gradients dan Hierarchial Centroid Untuk Sketch Based Image Retrieval *Jurnal Teknik ITS* **5** 311-316
[6] Haralick R.M, K. Shanmugan, I. Dinstein Textural Features for Image Classification *IEEE Transactions on Systems, Man, and Cybernetics* **6** 610-621
[7] Lletı R, Ortiz MC, Sarabia LA, Sánchez MS Selecting variables for k-means cluster analysis by using a genetic algorithm thar optimises the silhouettes *Analytica Chimica Acta* **515** 87-100