Sketch-Based Image Retrieval with Histogram of Oriented Gradients and Hierarchical Centroid Methods

Viny Christanti Mawardi[1,*], *Yoferen* Yoferen[1], and *Stéphane* Bressan[2]

[1]Faculty of Information Technology, Tarumanagara University, Jl. Letjen. S. Parman no. 1, Jakarta 11440, Indonesia
[2]School of Computing, National University of Singapore (NUS),
13 Computing Drive, 117417 Singapore

Abstract. Searching images from digital image dataset can be done using sketch-based image retrieval that performs retrieval based on the similarity between dataset images and sketch image input. Preprocessing is done by using Canny Edge Detection to detect edges of dataset images. Feature extraction will be done using Histogram of Oriented Gradients and Hierarchical Centroid on the sketch image and all the preprocessed dataset images. The features distance between sketch image and all dataset images is calculated by Euclidean Distance. Dataset images used in the test consist of 10 classes. The test results show Histogram of Oriented Gradients, Hierarchical Centroid, and combination of both methods with low and high threshold of 0.05 and 0.5 have average precision and recall values of 90.8 %, and 13.45 %, 70 % and 10.64 %, 91.4 % and 13.58 %. The average precision and recall values with low and high threshold of 0.01 and 0.1, 0.3 and 0.7 are 87.2 % and 13.19 %, 86.7 % and 12.57 %. Combination of the Histogram of Oriented Gradients and Hierarchical Centroid methods with low and high threshold of 0.05 and 0.5 produce better retrieval results than using the method individually or using other low and high threshold.

Keywords: Canny edge, content-based image retrieval, dataset image, digital image processing, sketch image.

1 Introduction

The technique used to perform digital image processing is image retrieval. Image retrieval is a computer system for searching and restoring images from a collection of digital images [1]. Image retrieval is usually done based on text, which is referred to as text-based image retrieval. The digital image collection will be given a keyword or description manually [2]. Thus, text-based image retrieval is very time-consuming to build, and only applies in one language.

Another technique that is often used is Content-Based Image Retrieval (CBIR). Content-based image retrieval is a method to search images by making comparisons between query image and all dataset images based on the information contained in that

---

[*]Correspondingauthor: viny@untar.ac.id

image. CBIR is one part of information retrieval which has now developed into Content-Based Multimedia Information Retrieval (CBMIR). CBMIR provides searching in various forms of media and by using various methods [3].

Content-based image retrieval focus to search information in the form of image data based on feature or characteristics of a given image. Image features can be in the form of: shape, colour, texture, and other forms, depending on the feature extraction method selected. Sketch-Based Image Retrieval (SBIR) is a development of the content-based image retrieval technique that performs image retrieval based on the similarity between each dataset images and the sketch image input [2]. This is a very useful thing in image retrieval, because it is easier to retrieve images by drawing the image you want to look for, compared to retrieve images by typing keyword for the image you want to look for, because each person can have different opinion on what keyword that fits the most.

The earlier method used for sketch-based image retrieval is Query by Visual Example (QVE) that uses edge maps as the sketch image. But this process has not been supported by a lot of software and requires a lot of funds [4]. In this research, the sketch-based image retrieval system designed using the Histogram of Oriented Gradients and Hierarchical Centroid methods. Pre-processing dataset images is done by the Canny Edge Detection method, and retrieval will be carried out based on differences in features between sketch images and each dataset images calculated using Euclidean Distance.

2 Research methods

This section explains definition, function, formula, and how to apply some of the methods used in order to build the sketch-based image retrieval.

2.1 Canny edge detection

Canny Edge Detection is the edge detection method to detect edges from the dataset RGB images. The steps for applying Canny Edge Detection are as follows [5]: (i), Gaussian Filter is used to blur images and reduce noises. (ii), Gradient Magnitude and Direction can be calculated by multiplying sobel x-direction and y-direction matrices with the image matrix. Multiplication is done by convolution. (iii), Non-maximum Suppression is used for thinning the edges of the image [6]. The calculation of non-maximum suppression is done for each pixel in the image, and checks whether the pixel value is a local maximum compared to their neighbor pixels. Non-maximum suppression is done by simplifying the gradient direction values into four categories of angles, namely 0, 45, 90, and 135.

Then, pixels are compared to their neighbour pixels based on their gradient direction. If the gradient magnitude of a pixel is greater than all the magnitudes of their neighbour pixels, then the pixel is categorized as a local maximum. Fourth, Hysteresis Thresholding uses two threshold values, namely high and low threshold. If the gradient magnitude is greater than the high threshold, then the pixel value is considered an edge and is assigned a value of 255. If the gradient magnitude is smaller than the low threshold, then the pixel value is not an edge and is given a value of 0. If the gradient magnitude is between the high and low threshold, the pixel value is considered an edge, when connected with edge pixels [7].

2.2 Histogram of oriented gradients

Histogram of Oriented Gradients (HOG) is a feature descriptor used in computer vision and image processing for object detection. Histogram of Oriented Gradients is window-based

descriptors that detect at points of interest. The extraction steps of the Histogram of Oriented Gradients feature are as follows: (i) Gradient Magnitude and Direction can be calculated by multiplying Dx and Dy matrices with the image matrix. Multiplication is done by convolution; (ii) Image Block Histogram is used to hold the gradient magnitude that is entered based on the gradient direction of the image. A histogram is created for each cell in the image. The size of a cell is 8 × 8 pixels. The histogram is divided into 9 bins which has a range from 0 to 180 degrees divided by the gradient direction of the image [8]; (iii) Feature Normalization is done for each image block. The size of an image block is 2 × 2 cell or 16 × 16 pixels [2]. Feature vectors of the Histogram of Oriented Gradients are obtained by doing normalization on the normalized block feature of the entire image.

2.3 Hierarchical Centroid

Hierarchical Centroid is a shape descriptor that is extracted recursively by decomposing image into sub-imagesbased on the tree decomposition. The length of the descriptor is $2 \times (2^d - 2)$ where d is the depth of the feature extraction for Hierarchical Centroid.

The Hierarchical Centroid feature can be extracted using the following algorithm [8]: (i) Take the input image I and calculate the transpose of I; (ii) Calculate *centroid C(x_c, y_c)* from the root level; (iii) Recursively, divide image using centroid ($x = x_c$ or $y = y_c$) until reaching the desired depth. At each level, the axis of the coordinates is switched; (iv) Combine features extracted from image I and transposed image I; (v) Normalize HC feature vectors into [-0.5, 0.5] range.

Calculation of the distance between features can be done using Euclidean Distance. Euclidean Distance is a method for calculating distance between two points. Euclidean Distance is used to calculate the square root of the difference between two vectors [9]. In sketch-based image retrieval, the Euclidean Distance will be calculated between feature of sketch image input and all dataset images, where the features of each image are arranged into a one-dimensional vector.

Finally, the accuracy of SBIR counts with precision and recall. The precision and recall have been used to evaluate the performance of Histogram of Oriented Gradients and Hierarchical Centroid methods to do sketch-based image retrieval. In this research, the precision and recall was calculated based on the first 10 images from the retrieval results.

3 Results and discussions

In this research, all images were collected as dataset and two different scenarios was used for testing. First, test the feature extraction methods; and second, test the canny edge detection low and high threshold.

3.1 Dataset images and sketch images for testing

Dataset images used for testing consist of 10 classes, namely bottle, duffle bag, fidget spinner, guitar, study lamp, bowl, umbrella, bicycle, hockey stick, and beanie hat. Each class consists of 100 images, the total dataset images is 1 000 images, and obtained from the internet. The example of real image and sketch image that already draw by user can be seen at Figure 1.
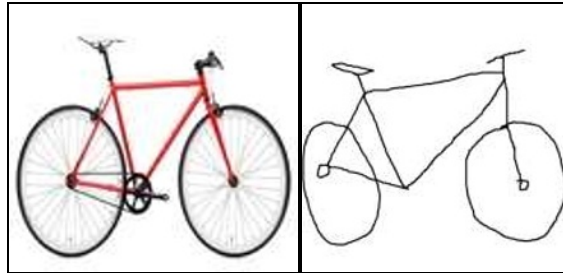
Fig. 1. Real image and sketch image example.

Sketch images used for testing are a drawn sketch of the image that you want to search. The number of retrieved images used in the test is 10 images, taken from the top 10 images from the retrieval results. Figure 2 is interface of the sketch based image retrieval that can try on the website. Figure 3 is the retrieval result after user draw umbrella sketch image.

Fig. 2. Interface of the sketch-based image retrieval (SBIR).

Fig. 3. Result of umbrella sketch-based retrieval.

3.2 Feature extraction methods testing

Feature extraction methods' testing is done by calculating the precision and recall values based on the retrieval results using a particular feature extraction method, with 100 sketch images from ten classes. Feature extraction methods tested are Histogram of Oriented Gradients, Hierarchical Centroid, and combination of both methods

Based on the test results at Table 1, it can be seen that the combination of the Histogram of Oriented Gradients and Hierarchical Centroid methods with 91.4 % and 13.58 % precision and recall value produce better retrieval results compared to other methods, because more features are compared. The Histogram of Oriented Gradients method with precision and recall of 90.8 % and 13.45 %, produce better retrieval results than the Hierarchical Centroid method with 70 % and 10.64 % precision and recall value. Because there are classes that have different shapes but with centroid points that are similar to other classes.

Table 1. Average precision and recall values from feature extraction methods testing with 100 sketch queries and 1 000 dataset images.

Feature Extraction Method	Precision	Recall
Histogram of Oriented Gradients	90.8 %	13.45 %
Hierarchical Centroid	70 %	10.64 %
Histogram of Oriented Gradients and Hierarchical Centroid	91.4 %	13.58 %

3.3 Canny edge detection low and high threshold testing

Low and high threshold testing is done by calculating the precision and recall values based on the retrieval results using a particular threshold with Histogram of Oriented Gradients and Hierarchical Centroid methods, with 100 sketch images from 10 classes. Low and high thresholds tested are 0.01 and 0.1, 0.05 and 0.5, and 0.3 and 0.7.

Table 2. Average precision and recall values from low threshold and high threshold testing with 100 sketch queries and 1 000 dataset images.

Low Threshold and High Threshold Canny Edge Detection	Precision (%)	Recall (%)
0.01 and 0.1	87.2	13.19
0.05 and 0.5	91.4	13.58
0.3 and 0.7	86.7	12.57

Based on the test results, at Table 2 can be seen that low and high thresholds of 0.05 and 0.5 with 91.4 % and 13.58 % precision and recall value produce better retrieval results compared to other thresholds, because with this threshold, edges produced represent the image well and does not contain too much noise. Low and high threshold 0.01 and 0.1 with precision and recall 87.2 % and 13.19 %, produce better retrieval results than low and high thresholds 0.3 and 0.7 with 86.7 % and 12.57 % precision and recall value. Because the edge results are detailed so images with different classes can be distinguished.

3.4 Sketch image testing

Sketch image testing is done by calculating the precision and recall of the retrieval results with the Histogram of Oriented Gradients and Hierarchical Centroid methods with a low and high threshold of 0.05 and 0.5, with 20 sketch images from 10 classes. The tests produce average precision and recall of 67 % and 9.90 %.

From Table 3, can be seen that duffle bag and fidget spinner image class with precision value of 80 % produce the best retrieval results compared to other dataset image class, because the class has a unique shape. The bottle and study lamp image classes with a precision value of 55 % produce the worse retrieval results compared to other dataset image class. For the bottle images, it is difficult to distinguish between long neck bottles and short neck bottles. For the study lamp images, the drawn query sketches are only curved that resemble question marks that have a similar shape to other classes than the study lamp, such as bottle classes and umbrella classes.

Table 3. Average precision and recall values from sketch images testing for each class with 20 sketch queries and 1 000 dataset images

Dataset Image Class	Precision (%)	Recall (%)
Bottle	55	11
Duffle Bag	80	8
Fidget Spinner	80	8
Guitar	60	12.18
Study Lamp	55	11.28
Bowl	65	6.5
Umbrella	65	13.03
Bicycle	60	6
Hockey Stick	75	15.11
Beanie Hat	75	7.5

4 Conclusions

The conclusions obtained after testing the sketch-based image retrieval program with Histogram of Oriented Gradients and Hierarchical Centroid methods are as follows: (i) The combination of Histogram of Oriented Gradients and Hierarchical Centroid methods produce better results compared to using the method individually. And Histogram of Oriented Gradients method produce better retrieval results compared to Hierarchical Centroid method; (ii) Canny Edge Detection preprocessing method with low and high threshold 0.05 and 0.5 produce better retrieval results compared to using low and high threshold 0.01 and 0.1 or 0.3 and 0.7, because the edge produced by the threshold represents the image well and does not contain too much noise; (iii) Fidget spinner image class produce the best retrieval results compared to other image classes, because fidget spinner image has a unique shape. The bottle image classes produce the worst retrieval results compared to other image classes, because bottle image has a similar shape to guitar image, so the search results sometimes do not match the query image.

References

1. S. Bansod, D.S. Gawande, R. Thakur. IJRITCC **5**,3:455–459(2017). https://ijritcc.org/index.php/ijritcc/article/view/328
2. A.F. Randa, N. Suciati, D.A. Navastara. Jurnal Teknik ITS. **5**,2:A311–A316(2016). https://media.neliti.com/media/publications/191789-ID-none.pdf
3. M.S. Lew, N. Sebe, C. Djeraba, R. Jain. ACM T. Multim. Comput. **2**,1:1–19(2006). https://dl.acm.org/doi/10.1145/1126004.1126005

4. N.M. Asiri, N. AlHumaidi, N. AlOsaim. *Combination of histogram of oriented gradients and hierarchical centroid for sketch-based image retrieval.* Paper Presented in The Second International Conference on Computer Science, Computer Engineering, and Social Media (CSCESM) (Lodz, Poland, 2015). IEEE 149–152(2015). https://ieeexplore.ieee.org/document/7331884
5. Susila. Majalah ilmiah INTI, **12**,2:235–240(2017). [in Bahasa Indonesia]. https://ejurnal.stmik-budidarma.ac.id/index.php/inti/article/view/430
6. J. Liang. *Canny edge detection*. [Online] from http://justin-liang.com/tutorials/canny/#suppression,%205%20September%202018 (2018). [Accessed on July 28, 2019]
7. A.L. Kabade, V.G. Sangam. IJARECE **5**,5:1292–1295(2016). http://ijarece.org/wp-content/uploads/2016/05/IJARECE-VOL-5-ISSUE-5-1292-1295.pdf
8. N. Dalal, B. Triggs. *Histograms of oriented gradients for human detection*. Paper Presented in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (San Diego, California, 2005). https://dl.acm.org/doi/10.1109/CVPR.2005.177
9. E. Ilunga-Mbuyamba, J.G. Avina-Cervantes, D. Lindner, J. Guerrero-Turrubiates, C. Chalopin. *Automatic brain tumor tissue detection based on hierarchical centroid shape descriptor in Tl-weighted MR images.* Paper Presented in 2016 International Conference on Electronics, Communications and Computers (CONIELECOMP) (Cholula Puebla, Mexico, 2016). IEEE, 62–67(2016). https://ieeexplore.ieee.org/document/7438553
10. S.R. Wurdianarto, S. Novianto, U. Rosyidah. Techno.Com **13**,1:31–37(2014). [in Bahasa Indonesia] http://publikasi.dinus.ac.id/index.php/technoc/article/view/539