Filtering Selfie Face Image on Instagram using Face Detection

Jewel Mahmud Nimul Shamim
Department of Computer Science
Memorial University of Newfoundland
NL, Canada
jmnshamim@mun.ca

***Abstract*— Instagram is one of the famous and fast-growing photos and videos sharing platform. Instagram also allows users to select filters to photos and share them with followers. However, people mostly use Instagram to share their selfie face photo, which is also dominated content. There are plenty of ways to search for images on Instagram, but one of the most familiar way is 'hashtag.' But there are no rules for hashtag; that is why most of the time, the hashtag does not match the uploaded image. The purpose of this study using the face detection technique, separate selfie face images on search results based on hashtags. Besides, in this study, we also do some image preprocessing techniques while detecting a human face and compare the result with the original paper. We use the Haar Cascade classifier to detect the selfie face, which is 64.6% accurate to detect the human face.**

Keywords—face detection, image preprocessing, selfie face, Instagram photo search

I. INTRODUCTION

Social media makes people spend more time on the internet rather than reading books. Social media are also created to share content instantly, efficiently, and in real-time [3]. Instagram is one of the successful photo-sharing social media platforms with the greatest number of users. Based on data from TechCrunch, as of June 2018, Instagram has touched the 1 billion monthly active user record [1], which has grown from 600 million active users in April 2017 [2].

Instagram is a visual platform, and the purpose is to allow users to share images and videos with their friends. Every day more than billions of photos are uploaded to Instagram, which is very abundant. In the Asia Pacific, Indonesia has the largest Instagram users, with a total of 45 million. Besides, it is a source for people to search for a reference for a tourist place, restaurant, or location in an area [2].

On Instagram, people upload photos with a keyword that is also called a hashtag. Instagram has an option where users can search for photos using specific keywords or hashtag[3]. But most of the cases, the uploaded photo and the hashtag not related because there are no specific rules for the hashtag and the uploaded image. The most important topics and the relationship among content topics of images connected with a piece of appropriate hashtag information are low [2].

The purpose of this study using the face detection technique, separate selfie face images on search results based on hashtags. Besides, in this study, we also do some image preprocessing techniques to detect low light selfie photos on Instagram. The model uses a web data extraction technique using the open-source library Instagram scrapper to download Instagram photos using the hashtag and face detection techniques applying the Haar Cascade method to detect the human face. [2]. Because of the simplicity and efficiency, we use the Haar Cascade classifier to detect a human face. Besides, the preprocessing technique also used to identify the dark and bright image using the mean (averaging) filter. Depending on the dark and bright image, we use the histogram equalization method to modify the dynamic range and the contrast of the image and use a 3x3 gaussian filter to reduce the noise by smoothing the image. The preprocessing technique helps to identify even low light photos on Instagram.

II. LITERATURE REVIEW

A. Human Face Detection

In recent years, there are lots of studies that discuss human face detection using the Haar Cascade Classifier. Adri Priadana1, Muhammad Habibi2 studies use web extraction using beautifulsoup to download Instagram images and detect selfie face with the help of the Haar Cascade Classifier. [2]. Besides, they used the OpenCV library to perform the Haar cascade classifier, which gives high accuracy and superior performance with a simple background image [4]. Nevertheless, the study did not use any preprocessing technique that is why the face detection does not work on low light photo.

III. METHODOLOGY

The proposed method consists of seven steps starting from pre-processing the images to face detection. Fig. 1: shows all the steps of the application.

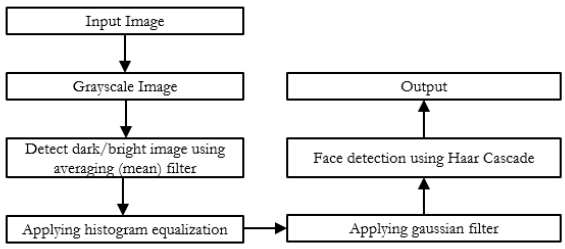

Figure. 1. Process of proposed system.

A. *Web Data Extraction*

The Instagram images consist of three fundamental parts, particularly: uploaded image, a caption, and hashtag. Hashtags are addressed with symbols #, which are used to index the subject on Instagram. This function enables users to search for the stuff that they are interested in. Fig. 2, demonstrate the basic anatomy of a search using a hashtag.

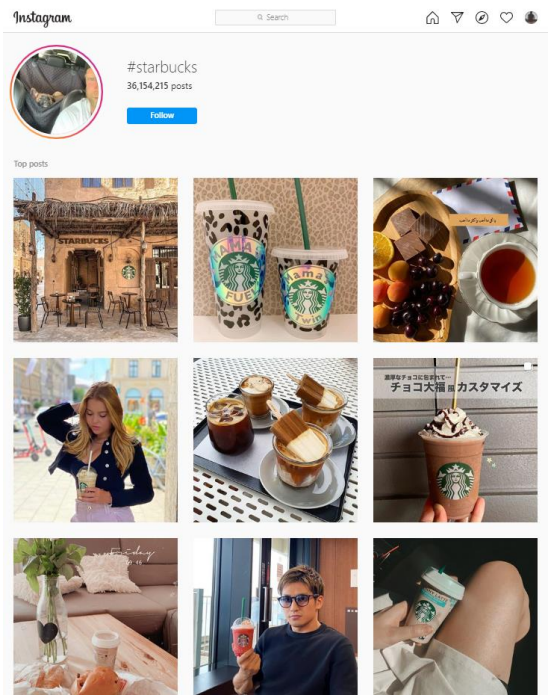

Fig. 2. An image search on Instagram based upon a hashtag.

The image downloading process uses the open-source python library Instagram-scrapper to extract image data from Instagram. Instagram-scrapper downloads the image with the given hashtag and stores it into a directory with the name of the hashtag.

B. *Image Pre-processing and detect dark and bright image*

Firstly, the application reads a downloaded image and convert images into grayscale. Afterward, the application separates dark and bright images. In this section, we blur the image with a 3x3 averaging (mean) filter and get the threshold value from the result. In the dark image, the threshold value will be less than 127, and for the bright image, it will vary between 127 to 255. We consider the image is dark if the threshold value is less than 50.

C. *Histogram Equalization*

The proposed method also detects selfie face image even the photo taken in low light. In this step, we took the detected low light image and performed the histogram equalization method using the OpenCV function cv2.equalizeHist. As the histogram equalization method use for modifying the dynamic range and the contract of an image. The histogram equalization operators implement monotonic nonlinear mapping. That is why the produced image contains a uniform combination of intensities.

D. *Gaussian Filter*

Afterward, a Gaussian filter with a 3x3 kernel uses to blur and reduce the noise on the image, which helps to detect the outer edge of the face. The Gaussian filter preserves more features like edges And in Instagram photos, its rarely visible noises (salt and pepper noise). That is the reason we are not using the median filter. Besides, as we are detecting a human face, we need to identify the outliner of the face.

E. *Haar Cascade Classifier*

In this project, we implement Haar Cascade Classifier to detect human faces with OpenCV. The Haar features are the central part and use to identify elements in the image. Haar feature uses the rectangle features to detect human face rapidly. The features are calculated by deducting the number of pixels in the black square to the number of pixels under the white square. The Haar cascade classifier uses the processed image and detect face in that image.

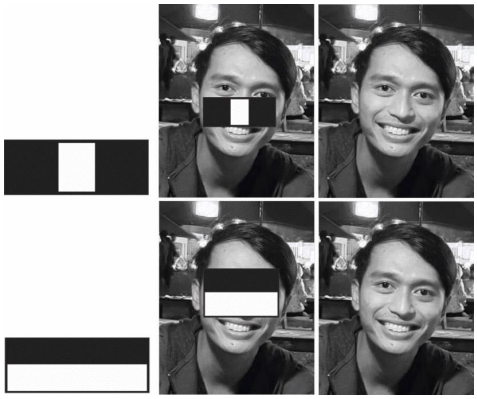

Fig. 3. Haar like features scan in images.[2]

F. *Human Face Detection in Instagram Images*

Lastly, the Haar cascade classifier method is applied to every single processed image. If it detects any human face on the image, the application draws a rectangle around the face. Also, the image will automatically be filtered and separate the photo into a different directory “Fig. 4”.

Fig. 3. Human face detection on Instagram photos.

IV. Results

The application in is project is developed with Python programming language, incorporated with the OpenCV on a desktop computer with AMD Ryzen 7 3800x 8-Core Processor @ 3.89 GHz CPU and the RAM is 32GB. It needs less than a minute to preprocess 1000 Instagram images and detect human faces.

The original paper did not apply any preprocessing technique to detect the human face. So, the application does not work for low light selfie photos “Fig. 4”. On the other hand, in this project, we first identify the dark image and follow a couple of image preprocessing techniques that produce an extraordinary result “Fig 5”.

Both examples 1 & 2 tested with two different low light photos. According to the original paper in both scenarios, the application cannot detect low light/dark photos. On the other side, the proposed application detects the human selfie faces after doing image preprocessing.

Fig. 4. (Example: 1) Original paper result

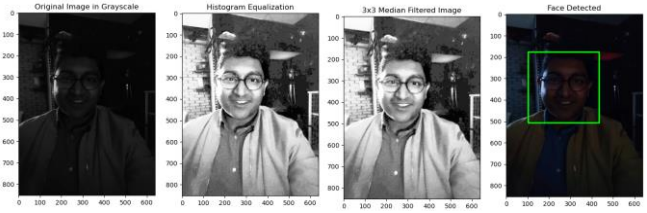

Fig. 5. (Example: 1) Proposed application result

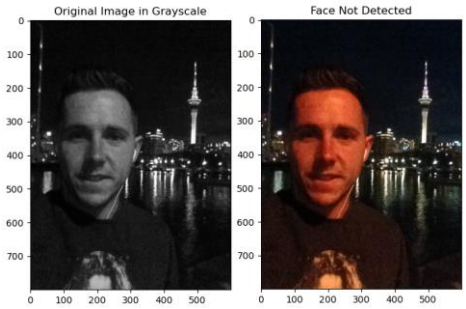

Fig. 6. (Example: 2) Original paper result

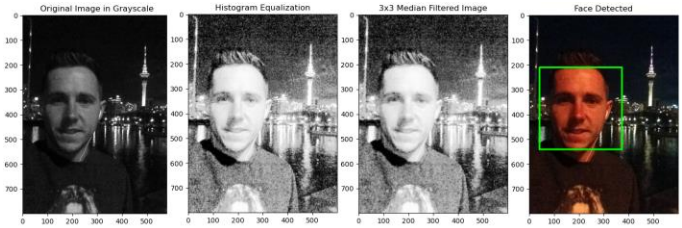

Fig. 7. (Example: 2) Proposed application result

V. Conclusion

In this project, we performed and analyzed Instagram images and detect human selfie face images using several hashtags. We have also shown how the Haar cascade classifier without image preprocessing cannot recognize low light/dark selfie images. On the other hand, using the image preprocessing technique enhances the process of identifying dark selfie images with the Haar cascade classifier with regular selfie face images. But as the Haar cascade accuracy value is 64,6%, so that is why in some cases, it faces difficulty in detecting the face.

VI. References

[1] "Techcrunch," 2019. [Online]. Available: https://www.wordstream.com/blog/ws/2017/04/20/instagram-statistics.

[2] A. Priadana and M. Habibi, "Face Detection using Haar Cascades," in *IEEE*, Yogyakarta, Indonesia, Indonesia, 2019.

[3] L. M. S. K. Yuheng Hu, "What we instagram: A first analysis of instagram photo content and user types," in *ICWSM*, Ann Arbor, United States, 2014.

[4] "TEMPO.CO," 26 06 2017. [Online]. Available: https://bisnis.tempo.co/read/894605/45-juta-pengguna-instagram-indonesia-pasar-terbesar-di-asia.

[5] "OpenCV," [Online]. Available: https://docs.opencv.org/master/d4/d13/tutorial_py_filtering.html.

[6] "OpenCV Haar Cascade Classifier," [Online]. Available: https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html.