Detection of Faces from Images Using Haar Cascade Classifier

THIN THIN SWE
Lecturer, Computer University (Myitkyina), Kachin State, Myanmar

Abstract- ***Nowadays, the increasing volume of images is absolutely demanded in most of digital image processing and pattern recognition. Moreover, face detection from images has become essential as it can be applied in various areas such as surveillance system, biometrics, gender classification, and so on. In this paper, Haar Cascade Classifier of Open Source Computer Vision Library (OpenCV) is utilized in detection of faces. In addition, Apache Hadoop, a distributed processing platform is applied to solve the computation time burden of face detection from large-scale images. According to the experimentation, the face detection with haar cascade classifier which is experimented on Apache Hadoop platform can offer satisfactory execution time results.***

Indexed Terms- Apache Hadoop, Face Detection, Haar Cascade Classifier, OpenCV

I. INTRODUCTION

Nowadays, a large collection of images and videos creates unstructured large-scale data in almost 80 percent of public and business corporations. In general, an image can be assumed as a two-dimensional data matrix which can be specifically arranged in rows and columns. Typically, every element of this matrix is called image element or pixel. As each pixel consists of its respective RGB values, high-resolution image can also be referred to high-dimensional data space [2], [4]. Face detection is one of the applications under computer vision technology. It is intended to find and identify human faces in digital images. Moreover, it is the process of locating faces or objects in images to perform as a preliminary step for face recognition. Face detection can be used in a variety of fields such as law enforcement, biometrics, and entertainment and so on to support surveillance and tracking of people in real time [1]. According to studies and research, automatic face detection is a complex problem in image processing and pattern recognition. In addition, haar cascade classifier is a machine learning-based approach for object detection. It is an effective object detection method where a cascade function is trained from a large number of positive and negative images. These images are then used to detect objects in other images. In order to obtain satisfactory detection, the classifiers are needed to train by using thousands to millions of images [3], [7].

Applying the haar cascade classifier for face detection process can actually offer simplicity and effectiveness. As for reliability and correctness in face detection from images, haar cascade classifier is one of the best detectors which can be utilized in real time applications [9]. However, face detection from large-scale or a large amount of images may take longer computation time. Therefore, a distributed processing platform is needed to implement parallel image processing scheme especially for face detection process. The paper is organized as follows: the related works which motivating the proposed work is briefly presented in section 2. The explanation about face detection process including haar features and haar cascade classifier is expressed in section 3. The proposed system is expressed in section 4 and the detailed explanation of implementation, experimentation, results and discussions about the proposed system is then described in section 5. Finally, the conclusion and future works are presented in section 6.

II. RELATED WORKS

Phillip Ian Wilson and Dr. John Fernandez [6] presented that although there are many diverse algorithms for face detection, each has its own strengths and weaknesses. They also discussed about haar cascade classifiers can be applied to accurately detect facial features in images and it is needed to train

in detecting facial features such as mouth, eyes, and nose. Vandna Singh, Dr. Vinod Shokeen and Bhupendra Singh [5] presented the issues and problems of face detection from images with simple and complex backgrounds. They also expressed that the haar cascade classifier can offer superior performance with simple background images. Zebin Wu and et al. [8] explored that increasing volume of hyper spectral images is absolutely required to consider in terms of storage and data processing. Thus, they suggested that large-scale hyper spectral image processing to be implemented on parallel and distributed platforms such as MapReduce and Spark. In their system, the implementation of PCA on Spark platform obtained better speedups when compared to traditional processing platform.

III. FACE DETECTION

Face detection can perform as a first and essential step for face recognition. It is the process of locating and determining the position of faces or objects in images by detection process. Typically, face detection can be divided into three stages. In first stage, the effect of interfering factors is reduced by using histogram equalization and noise reduction. The second stage determines the regions with high possibility where a face can be placed. In third or final stage, verification process is performed on previously selected regions to detect and mark the output face [7]. The most significant face detection methods can be categorized into:

- Knowledge based methods
- Featured-based methods
- Template matching

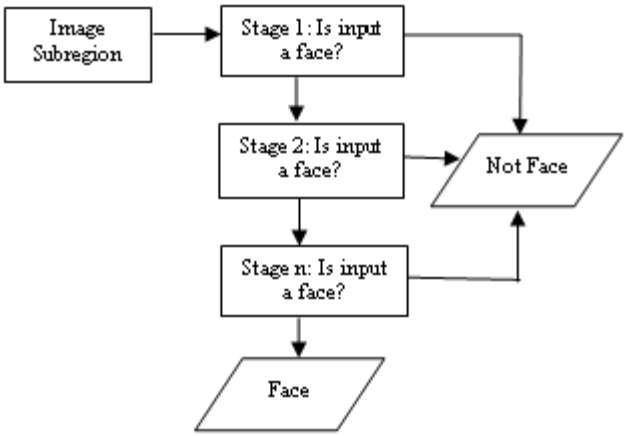

Fig.1 The general working flow of face detection

A. Face Detection using Haar Cascade Classifier

Face detection typically uses classifiers which are algorithms to detect whether it is a face or not a face in an image. For detecting faces, these classifiers have to be trained by using thousands to millions of images in order to obtain better accuracy. In addition, Open Source Computer Vision Library (OpenCV), a library of programming functions, which is intended to utilize in real-time computer vision applications. It uses two kinds of classifiers:

- LBP (Local Binary Pattern)
- Haar Cascades

In fact, haar cascade classifier is based on haar wavelet technique to analyze pixels in the image into squares by function. It commonly uses machine learning models to obtain good accuracy from training data. By using training data, the classifier identifies features which can be considered as a face. Besides, it utilizes integral image concepts for computing features detected. Haar cascades also apply the Adaboost machine learning algorithm which selects a small number of important features from a large set to offer an efficient detection result of classifier [5], [6].

B. Haar Features

For human face detection, haar features are the most important part of the classifier. Haar features are applied to detect the existence of features in the given image. Each feature typically produces a single value. It is calculated by subtracting the number of pixels via the white rectangle from the number of pixels via the black rectangle. Haar features exist in the form of

rectangle features for face detection which are shown in Figure 2.

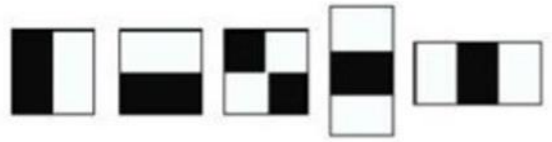

Fig.2 Haar features

IV. THE PROPOSED SYSTEM

The overall architecture of the proposed system is shown in Figure 3.

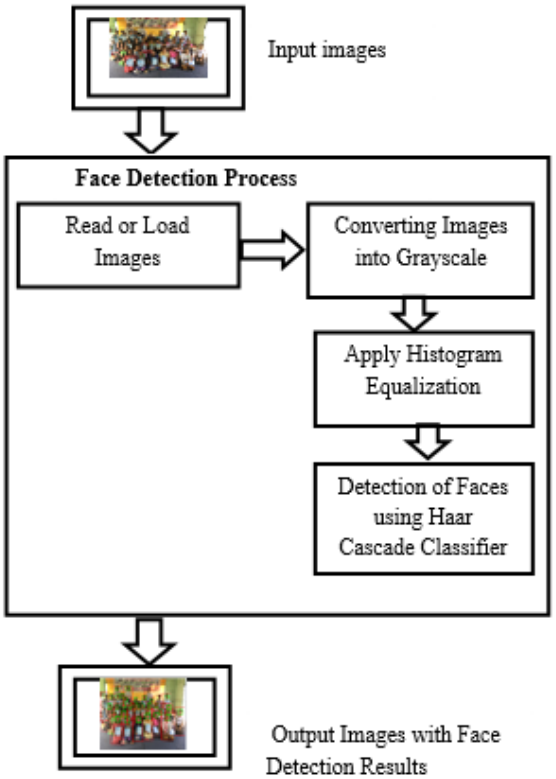

Fig. 3 The Proposed System

Firstly, the images are collected to be processed from various source of images such as ImageNet, Kaggle, peopleimages [10], [11], [12]. These images are taken as input images to face detection process. Open Source Computer Vision Library (OpenCV) is applied to implement the haar cascade classifier in this system. Firstly, loading the images and transforming them into grayscale are performed. The purpose for transforming gray channel is that easy to process and computationally less intensive. Then, histogram equalization is utilized on images. Finally, OpenCV's haar cascade classifier step-by-step detects faces in images for face detection results.

Fig. 4 Face detection result of a sample image

V. EXPERIMENTATION

The face detection process of the system is experimented on two processing platforms: traditional processing platform (standalone computer) and distributed processing platform (cluster of computers).

A. Apache Hadoop

Apache Hadoop, a distributed processing platform, provides highly parallelizable and executable user desired programs or applications on a cluster of computers [8]. It follows the design of master-slave architecture supporting massive storage for any types of data.

B. Experimental Setup

To implement face detection process on distributed platform, the experiments are performed on a cluster of four computers, one for "Master" (Server node) and three "Slave" nodes. The cluster runs Linux Red Hat 4.6.3, and Apache Hadoop Distribution 2.8.3 were already installed on this cluster.

C. Results and Discussions

The original input images are taken from ImageNet, Kaggle, peopleimages [10], [11], [12] as large-scale image collection for the system. These images are sent to the face detection process. According to the experimentation, the image collection including more human faces is potential to offer wrong detection results, for example, image collection_3 has encountered four hundreds incorrectly detected faces and image collection_2 gives two hundreds incorrect

detection results. Furthermore, the execution time of face detection using traditional processing platform are also described in Table 1. According to the Table 1, image collection_1 enables to detect faces completely with 8800 seconds (147 minutes). However, it is not satisfactory result for large-scale images. Thus, the proposed system is developed on Apache Hadoop for face detection from large-scale images. In Table 2, the face detection process which is implemented on Apache Hadoop can efficiently minimize the computation or execution time. For example, in detecting 3700 faces in image collection _3 takes 12600 seconds instead of 22900 seconds in traditional platform. It is observed that although the detection performance in two platforms is the same, the execution time resulted from distributed platform is faster than traditional platform. Therefore, face detection on distributed platform especially Apache Hadoop should be chosen for large-scale image collection containing a large number of images.

Table I. Face detection on traditional platform

Image collection	Total no. of faces	No. of correctly detected faces	No. of incorrectly detected faces	Running Time (Seconds)
1	1000	850	150	8800
2	3300	3100	200	20400
3	3700	3300	400	22900
4	500	500	0	4770
5	200	200	0	2150

Table II. Face detection on Apache Hadoop platform

Image collection	Total no. of faces	No. of correctly detected faces	No. of incorrectly detected faces	Running Time (Seconds)
1	1000	850	150	4500
2	3300	3100	200	10400
3	3700	3300	400	12600
4	500	500	0	2130
5	200	200	0	950

CONCLUSION

Face detection is a first and essential step for face recognition intended for localizing and extracting the face region from image. In this system, OpenCV is used to implement the haar cascade classifier. Furthermore, Apache Hadoop, a distributed processing platform is applied to solve the computation time burden of face detection from large-scale images. According to the experimentation, although the detection performance between traditional and distributed platform is the same, the execution time resulted from distributed platform through Apache Hadoop platform is faster than traditional one. In future works, it is intended to extend the proposed system with the purpose of classifying different facial expressions for subsequent recognition process.

REFERENCES

[1] B. K. Gunturk, et al., -Eigenface-based super-resolution for face recognition,| *in Proc. International Conference on Image Processing, IEEE,* 2002.

[2] B. Mohammed Amin, S. Mahmoudi, and B. Ghalem, -PCA as dimensionality reduction for large-scale image retrieval systems,| International *Journal of Ambient Computing and Intelligence (IJACI)*, vol. 8, pp. 45-58, 2017.

[3] G. Ajit P, S. R. Khot, -Facial expression recognition using principal component analysis,| *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 3, pp. 2231-2307, 2013.

[4] Ng. S. C, -Principal component analysis to reduce dimension on digital image,| *in Proc. Procedia computer science 111*, 2017, pp. 113-119.

[5] S. Vinod, S. Bhupendra, Singh, -Face detection by haar cascade classifier with simple and complex backgrounds images using opencv implementation,| *International Journal of Advanced Technology in Engineering and Science*, vol. 1, pp.33-38, 2013.

[6] W. Phillip Ian, F. John, -Facial feature detection using Haar classifiers,| *Journal of Computing Sciences in Colleges*, vol. 21, pp. 127-133, 2006.

[7] Y. Ming-Hsuan, et al., -Detecting faces in images: A survey,| *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, pp. 34-58, 2002.

[8] Z. Wu, et al., -Parallel and distributed dimensionality reduction of hyperspectral data on cloud computing architectures,| *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, pp. 2270-2278, June 2016.

[9] Z. Zulhadi, S. Shahrel A. Suandi, -Face detection using combination of Neural Network and Adaboost,| *in Proc. TENCON 2011-2011 IEEE Region 10 Conference*, 2011.

[10] www.peopleimages.com

[11] www.image-net.org

[12] www.kaggle.com