International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-7, Issue-3, May-2019
DOI: 10.21276/ijircst.2019.7.3.8

# Image Search Engine and Individual Profile Building

**Shravya G, Prof. Smitha G R**

***Abstract*— Look technique utilized for the content gives semantically significant outcome, however isn't a similar with regards to the scan strategy utilized for pictures. Interactive media information is being distributed on the Web at an extraordinary rate. Likewise, in this time of innovation, it is conceivable to get data about any person from web. It has turned out to be fundamental to perform picture hunt of a person to recover the comparative pictures from Web. It is even conceivable to get any kind of data about any superstar from Wikipedia and different locales. This project aims at building the Image Search Engine for recovering the pictures just as structure the profile of a person, from World Wide Web. This is finished via preparing set of pictures of an individual and after that the web crawler creeps over the connections for getting the pertinent pictures. These recovered pictures coordinate with the name entered by the client. A similar outcome is utilized to get the data and manufacture the profile of a similar individual by slithering over the connections.**

***Index Terms: World Wide Web, Search Engine, Wikipedia***

## I. INTRODUCTION

The unstable development of the World Wide Web has turned out to be a twofold edged sword. While a huge measure of material is currently effectively open on the Web, finding explicit data remains a troublesome undertaking. An unpracticed client may discover it alongside difficult to discover the data required; even an accomplished client may miss significant Web pages [1]. Clients looking through the World Wide Web have various choices by and by accessible to them. Lycos, Excite, Alta Vista, and Yahoo! are nevertheless a couple of instances of valuable web indexes. These frameworks have been structured fundamentally to discover content put together data with respect to the Web. This undertaking is intended to locate the other real wellspring of data at present accessible on the Web: pictures [6].

**Manuscript received May 15, 2019**

**Shravya G**, M.Tech., Software Engineering, RV College of Engineering®, Bengaluru -59. (email: shravyag.sse17@rvce.edu.in**)**

**Prof. Smitha G R,** Assistant Professor, RV College of Engineering®, Bengaluru-59,

As the decent variety and size of advanced picture accumulations develop exponentially, productive picture recovery is ending up progressively essential. By and large, current programmed picture recovery frameworks can be portrayed into two classifications: content based and picture content-based. For content based picture recovery, the pictures are first commented on by content and afterward the content based Database Management Systems are utilized to perform picture recovery [7]. In this system, manual picture explanation is amazingly relentless and the visual substance of pictures is hard to be portrayed correctly by a constrained arrangement of content terms. To defeat these troubles, content-based picture recovery frameworks list pictures by their visual substance, for example, shading, shape, surface, and so on.

Most of the visual content based image retrieval systems are based on image database. The image resource in the database is limited and updated slowly. With the development of Internet technology, the fast-growing World Wide Web has become one of the most important sources of visual information. Efficient tools are needed to retrieve images from the Web [7]. Comparing to image databases, the Web is an unlimited, immense repository of images, covering much broader resources, and is increasing at an astonishing speed continuously. However, the Web is also a completely open information system without a well-defined structure. Image retrieval from the World Wide Web has to overcome great difficulties concerning speed, storage, computational cost, and retrieval quality [2].

The greater part of the visual substance based picture recovery frameworks depend on picture database. The picture asset in the database is restricted and refreshed gradually. With the advancement of Internet innovation, the quickly developing World Wide Web has turned out to be a standout amongst the most essential wellsprings of visual data [3]. Proficient apparatuses are expected to recover pictures from the Web. Contrasting with picture databases, the Web is a boundless, massive store of pictures, covering lot more extensive assets, and is expanding at an amazing velocity constantly [4]. Be that as it may, the Web is furthermore a totally open data framework without a very much characterized structure. Picture recovery from the World Wide Web needs to beat extraordinary challenges concerning speed, stockpiling, computational expense, and recovery quality. Some substance based picture web crawlers use Web crawlers to persistently cross the Internet, gather pictures, and concentrate highlights from the pictures [5]. Nonetheless, given the boundless information measure, the interest on computational power, picture transmission cost, and picture

 

stockpiling rapidly turns into a bottleneck for these frameworks [8].

Retrieving the images from Web will be insufficient. Web Crawlers crawls across the links in the Internet, fetches information about the image retrieved and thus builds the profile [9].

## II. PROBLEM STATEMENT

There are ordinarily when an individual need to think about somebody whose name the individual must have either overlooked or never known about. A similar circumstance can occur in Police Department who should need to discover past violations from a well-known face, or a Director who should need to cast a craftsman who is reasonable for a Play, unbeknownst the name.

All these and comparable sorts have a typical issue, they've a picture of the individual, yet don't have a clue about the name or spot they may have seen that individual. Relatively few web indexes have choices to give a picture itself as a hunt question.

This task manages precisely such issues, with the goal that the client essentially needs to embed a picture to think about and run it, the outcome will be comparative pictures from the net and few sentences on the individual it perceives as to assemble a little profile on the perceived face from the picture.

## III. SYSTEM ARCHITECTURE

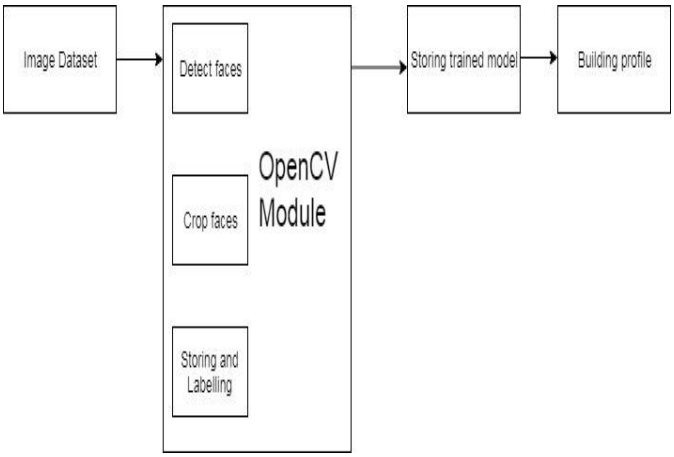


Fig. 1: System Architecture

Figure 1 shows the system architecture for the Image Search Engine and Individual Profile Building. The image dataset is the input by the user which undergoes pre-processing. This is done by OpenCV Module. The different steps in this are: first detecting the faces, cropping the faces and Storing and Labelling them. This labelled trained model is stored and the profile is built.

## IV. OVERVIEW OF METHODOLOGY

The project involves two different stages, Training and Profile Building. Set of pre-process images belonging to an individual are input by the user [10]. These images can be of any resolution or can have multiple faces of individuals as well. Accessing of individual images from the set of images input by the user and detection of individual faces takes place during pre-processing. Pre-processing step also involves cropping out just the faces that are detected from the images. These face-images are then stored by renaming them to labels that are associated with the name of the person. Training is then performed on the faces that are obtained from the directory of images. Output from the training process, that is all the data obtained from it is stored for future use [12].

Next step is the Profile Building task. The input and pre-processing steps from Training are performed. The output obtained will be just the faces with no unwanted data [11]. The face detected is then recognized using the training data it created and searches for a label that it predicts and then accesses the name. Once the face is recognized, it searches websites for related images belonging to that individual and scrapes them one by one. It then fetches for information on the same individual [13]. The images and information scraped from the internet are displayed one by one and hence the profile is built.

Therefore, this project involves two major objectives, to retrieve user intended images based on the input given by the user using Web Crawling mechanism and building the profile of the same individual after the retrieval of images [14].

## V. FACE DETECTING CLASSIFIERS

Classifier is a computer program which decides whether and image is a Face-image (Positive Image) or a Non-face image (Negative image). This classifier is trained on multiple images that include both positive and negative images that helps to learn how new image will be classified. OpenCV provides two different face detection classifiers: Haar Classifier and LBP Classifier. These two classifiers are used to process the images in gray scale. Since both the classifiers are pre-trained in OpenCV, all the necessary files are bundled together which will be present in /opencv/data. These files are loaded whenever any of the classifiers need to be run [1].

Haar classifier is based on machine-learning approach. It consists of cascade function which will be trained from multiple face and non-face images that finally help in detecting the objects (faces) in different images [1].

Image Search Engine and Individual Profsile Building is an application which undergoes through several processes to give the output. The image dataset input is first given by the user. This dataset undergoes pre-processing. The pre-processed images are then trained. This creates a trained model and gives each of the trained images a label. After the training of the images, the test image is recognized which does not belong to the dataset.

This identifies who the person is and fetches the information about the same individual from Web.

## VI. RESULTS AND DISCUSSION

Image Search Engine and Individual Profile Building is an application which undergoes through several processes to give the output. The image dataset input is first given by the user. This dataset undergoes pre-processing. The pre-processed images are then trained. This creates a trained model and gives each of the trained images a label. After the training of the images, the test image is recognized which does

 

ISSN: 2347-5552, Volume-7, Issue-3, May-2019
DOI: 10.21276/ijircst.2019.7.3.8

not belong to the dataset. This identifies who the person is and fetches the information about the same individual from Web.

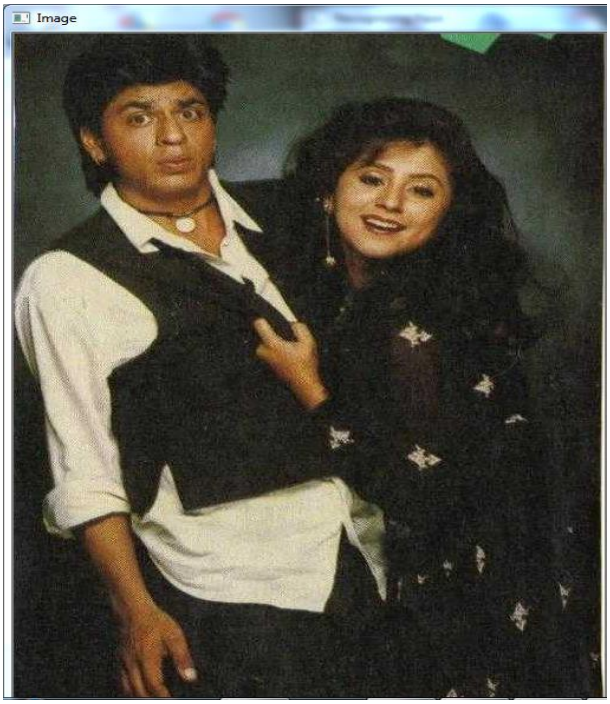


Fig. 2: Screenshot showing retrieval of images

Figure 2 shows the screenshot of how the images are retrieved from Web after saving the trained model.

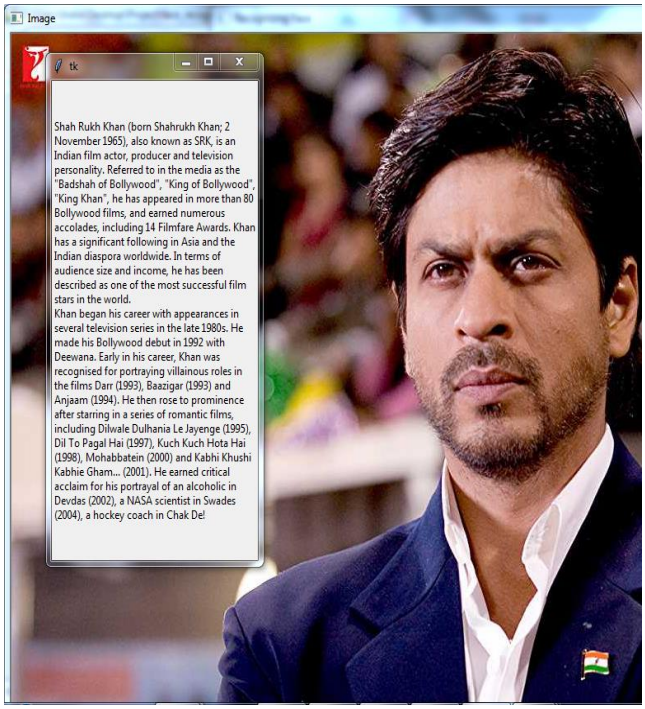


Fig. 3: Screenshot showing profile of an individual identified

Figure 3 shows how profile of an individual which is built after the retrieval of images from Web.

## VII. CONCLUSION

This project, by dealing with a unique problem, tries to solve it successfully using Python. OpenCV in python acts as a great tool for solving problems exactly like these despite its various other uses, ranging from cropping faces from images to recognizing them. Python further eases the process by its easy to understand coding structure. The dataset involving various images is pre-processed and used to train the model, further the model is saved for faster access during the consequent executions. After the training is over, test image is taken as input, face is detected. It is then recognized and label is found to further extract the name. The name obtained is used to scrape the web pages like IMDb and Wikipedia for retrieving images and information respectively. Thus, the profile is built and is shown to users as output.

## VIII. FUTURE SCOPE

Currently, OpenCV's Haar and LBHP models are used for face detection and recognition. In future, any stand-alone algorithm can be developed for more accurate and faster execution.

## ACKNOWLEDGMENT

This project work is supported as part of major project for M.Tech, Software Engineering Post Graduate Program in the R.V.College of Engineering®, Bengaluru-560059.

## REFERENCES

[1] Li Cuimei, Qi Zhilliang, Jia Nan, Wu Jianhua, "Human face detection algorithm via Haar cascade classifier combined with three additional classifiers," 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), 2017.

[2] Surbhi Jain, Joydip Dhar, "Image Based Search Engine Using Deep Learning," Proceedings of 2017 Tenth International Conference on Contemporary Computing (IC3), 10-12 August 2017, Noida, India.

[3] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, "Neural codes for image retrieval," in European conference on computer vision. Springer, 2017, pp. 584–599.

[4] R. Xia, Y. Pan, H. Lai, C. Liu, and S. Yan, "Supervised hashing for image retrieval via image representation learning," in the 1st International workshop on Edge Systems, Analytics and Networking, vol. 1, 2017, p. 2.10.1145/3213344.3212248/Vol 25 June 2017.

[5] K. Lin, H.-F. Yang, J.-H. Hsiao, and C.-S. Chen, "Deep learning of binary hash codes for fast image retrieval," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 27–35.

[6] N. Khosla and V. Venkataraman, "Building image-based shoe search using convolutional neural networks," CS231n Course Project Reports, 2017.

 

[7] Dhananjay Uttarwar, Aakash Agarwal, Riyaz Kadiwar and Vijay D. Katkar, "Distributed Content Based Image Search Engine using Hadoop Framework", International Conference on Communication and Signal Processing, April 6-8, 2017, India.

[8] Lu Zhang, Zhan Bu, Zhiang Wu and Jie Cao," DGWC: Distributed and Generic Web Crawler for Online Information Extraction", 2016 International Conference on Behavioral, Economic and Socio-cultural Computing (BESC), Vol. 0, No. 0, Article 00, Publication date: 2017 DOI: 10.1145/3057266.

[9] Wichian Premchaiswadi, Anucha Tungkatsathan, and SarayutIntarasema, "Improving Performance of Content-Based ImageRetrieval Schemes using Hadoop MapReduce", 2013 IEEE Biomedical Engineering, vol. 57, 2016.

[10] A.Bhagyalakshmi and V.VijayaChamundeeswari, "Image retrieval using Color and Texture Binary Patterns", 2016 International Conference on Green Computing and Internet of Things (ICGCIoT).

[11] Akash K Sabarad, Mohamed HumairKankudti, Meena S M, and Moula Husain, "Color and Texture Feature Extraction using Apache Hadoop Framework", 2016 International Conference on Computing Communication Control and Automation.

[12] David Edmundson and Gerald Schaefer, M. Emre Celebi, "Similarity-based Browsing of Image Search Results", 2016 IEEE International Symposium on Multimedia, Vol-7 (2.7) (2018) 335-340.

[13] R. Jenke, A. Peer and M. Buss, "Feature Extraction and Selection for Emotion Recognition from EEG," IEEE Transactions on Affective Computing, vol. 5, pp. 327-339, 2016.

[14] W. Plant and G. Schaefer, "Visualisation and browsing of image databases," in Multimedia Analysis, Processing and Communications, ser. Studies in Computational Intelligence. Springer, 2011, vol. 346, pp. 3–57M. Robnik- Š ikonja and I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF," Machine Learning, vol. 53, pp. 23-69, 2016.

 