Volume 3, Issue 8, August 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Introducing Celebrities in an Images Using HAAR Cascade Algorithm

Ms. Jaya M. Jadhav, Ms. Deipali V. Gore
Computer Engineering, Modern COE, Pune
Pune, India

Abstract— ***Now a day there is explosive growth in number of images available on the web. Among the different images celebrities images are also available in large amount which are in the form of posters, photographs and images taken at different events. Celebrities related queries are ranking high. Most of the end users are more interested in celebrity's related data and images. To better serve the end user demand we are going to develop an application which will provide celebrities information when an image is given as input.***

Keywords— Face detection methods, Common HAAR feature, HAAR Cascade Classifier, CFW dataset, Face Detection Algorithm.

I. Introduction

General web image page not always contain the name of celebrity in an image. Because of noise in web data it becomes difficult to identify celebrity name form web page text. There are mainly two challenges. Firstly the surround text of web image is lacking of standard grammar structure, therefore it is difficult to apply natural language processing techniques to extract celebrity names from it. Secondly the celebrities face in image may be having different pose, makeup, expression and occlusion caused by sunglasses or fancy hairstyles. So it becomes difficult to identify celebrity in an image with visual analysis and a normal face database. To face this challenge a CFW dataset can be used which contain millions of celebrity images in different pose, makeup, expression. Work in [1]–[4] were conducted on news images, where descriptive captions are usually provided and most of the time the caption contains the celebrity name that is there in an image. This paper presents different face detection approaches in section II; introduce HAAR Cascade Classifier in section III and CFW dataset in section IV.

II. Face Detection Methods

In an image annotation system first we need to detect the face in an image. Face detection methods are as follows. They are divided into four categories. [5] These categories may overlap, so an algorithm could belong to two or more categories. This classification can be made as follows:

- Knowledge-based methods or Ruled-based methods: that encodes our knowledge of human faces using different rules.
- Feature-invariant methods: Algorithms that try to find invariant features of a face despite its angle or position.
- Template matching methods: These algorithms compare input images with stored patterns of faces or features.
- Appearance-based methods: A template matching method whose pattern database is learnt from a set of training images.

Let us examine them in detail.

A. Knowledge - Based Methods

These are rule-based methods. They are based on our knowledge of faces, and translate them into a set of rules. It's easy to guess some simple rules. For example, a face usually has two symmetric eyes, and the eye area is darker than the cheeks. Nose is at the centre of face. Mouth is at the bottom of face area.

The big problem with these methods is that it is difficult to form appropriate set of rules. There could be high false positives rate if the rules were too general. On the other hand, there could be high false negatives rate if the rules were too detailed. A solution is to build hierarchical knowledge-based methods to overcome these problems. However, this approach alone is very limited. It's unable to find many faces in a complex image.

B. Feature-Invariant Methods

To overcome the limitations knowledge based methods for face detection. Feature-invariant methods are introduced. The method is divided in several steps. Firstly, it tries to find eye-analogue pixels, so it removes unwanted pixels from the image. After performing the segmentation process, they consider each eye-analogue segment as a candidate of one of the eyes. Then, a set of rule is executed to determinate the potential pair of eyes. Once the eyes are selected, the algorithm calculates the face area as a rectangle. The four vertexes of the face are determined by a set of functions. Finally, they apply a cost function to make the final selection. They report a success rate of 94%, even in photographs with many faces. These methods are efficient with simple inputs. But, there is problem to detect eye pixels if a man is wearing glasses? In that case there are other features used for face detection. For example, there are algorithms that detect face-like textures or the color of human skin.

There are different colour models. We can use more than one colour model for better face detection. For example, RGB and

HSV are used together successfully.

$$0.4 \leq r \leq 0.6, 0.22 \leq g \leq 0.33, r > g > (1-r)/2$$
$$0 \leq H \leq 0.2, 0.3 \leq S \leq 0.7, 0.22 \leq V \leq 0.8$$

Both conditions are used to detect skin colour pixels. However, these methods also have limitations. Skin colour can vary significantly if light conditions change. All over the world we have people with very dark skin colour and some may have very bright skin colour.

C. Template Matching Methods

These methods define a face as a function. We try to find a standard template of all the faces. Two templates are discussed here. First, a face can be divided into eyes, face contour, nose and mouth. Also a face model can be built by edges. But these methods are limited to faces that are frontal and unclouded. Other templates use the relation between face regions in terms of brightness and darkness. These standard patterns are compared to the input images to detect faces. This approach is simple to implement, but it's inadequate for face detection. It cannot achieve good results with variations in pose, scale and shape.

D. Appearance-Based Methods

The templates in appearance-based methods are learned from the examples in the images. In general, appearance-based methods rely on techniques from statistical analysis and machine learning to find the relevant characteristics of face images. Some appearance-based methods work in a probabilistic network. An image or feature vector is a random variable with some probability of belonging to a face or not. Another approach is to define a function between face and non-face classes. These methods are also used in feature extraction for face recognition.

Above methods can be summarized in the following Table I.

TABLE I

FACE DETECTION APPROACHES WITH STRENGTH AND LIMITATIONS

Method	**Strengths**	**Limitations**
Knowledge - based methods	It's easy to guess some simple rules	Difficulty in building an appropriate set of rules. It's unable to find many faces in a
Feature-invariant methods	Success rate of 94%. If Face is with sunglasses, Skin colour detect the face	Skin colour can vary significantly if light conditions change.
Template matching methods	define a face as a function simple to implement	Limited to faces that are frontal, cannot achieve good results with variations in pose, scale and
Appearance-based methods	Use a wide variety of classification methods. Sometimes two or more classifiers are combined to achieve better	Rely on techniques from statistical analysis and machine learning to find relevant characteristics of face images.

III. FACIAL FEATURE DETECTION USING HAAR CLASSIFIERS

Viola and Jones [6] introduced a method for accurate and rapid face detection within an image. This technique accurately detects facial features, because the area of the image being analysed for a facial feature needs to be regionalized to the location with the highest probability of containing the feature. For example eyes can detected at the upper part of the face, mouth is at bottom, nose is at the centre of face. By regionalizing the detection area, false positives are eliminated and the speed of detection is increased due to the reduction of the area examined. Many different algorithms exist to perform face detection; each has some strengths and limitations. Most of them are based on analysis of pixels. These algorithms suffer from the same problem; they are slow and expensive. Any image is only a collection of colour and/or light intensity values. Analysing these pixels for face detection is lengthy process and also difficult to accomplish because of the wide variations of shape and pigmentation within a human face. So Viola and Jones devised an algorithm, called HAAR Classifiers, to rapidly detect any object, including human faces, using AdaBoost classifier cascades that are based on HAAR-like features and not pixels.

A. HAAR Cascade Classifiers

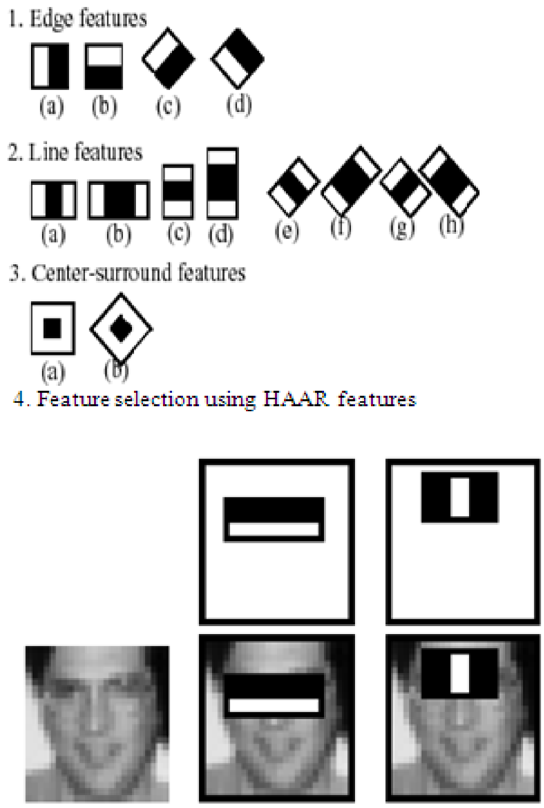

Fig. 1 Common HAAR Features

The important factor for HAAR classifier object detection is HAAR - like feature. These features are based on change in Contrast values between adjacent rectangular groups of pixels rather than the intensity values of a single pixel. For a human face, eye area pixels are dark than the nose area pixels. The contrast values between adjacent rectangular groups of pixels are used to determine relative light and dark areas. Two or three adjacent groups with a relative contrast groups form a HAAR-like feature. HAAR like features are as shown in Fig. 1 that is used to detect an image. Fig. 1 also shows how these features detect eye and nose for a face image.

B. Integral Image

Integral image is an intermediate representation of an image. It is used to calculate the simple rectangular features of an image. The integral image is an array containing the sums of the pixels' intensity values located directly to the left of a pixel and directly above the pixel at location (x, y) inclusive. So AI[x, y] is the integral image for the original image A[x, y] and it is computed as shown in equation 1 and illustrated in Fig 2.

$$AI[x,y] = \sum_{x' \le x, y' \le y} A[x',y'] \qquad (1)$$

$$AR[x,y] = \sum_{x' \le x, x' \le x-|y-y'|} A(x',y') \qquad (2)$$

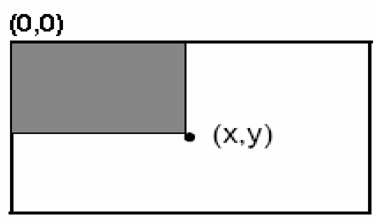

Fig. 2 Summed area of integral Image

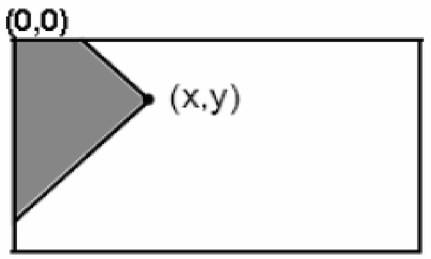

Fig. 3 Summed area of rotated integral image

Lienhart and Maydt introduced the features rotated by forty-five degrees, like the line feature shown in Fig. 1(e). These require another intermediate representation called the rotated integral image or rotated sum auxiliary image. The rotated integral image is calculated by finding the sum of the pixels' intensity values that are located at a forty five degree angle to the left and above for the x value and below for the y value. So AR[x, y] is the rotated integral image for the original image A[x, y] and it is computed as shown in equation (2) and illustrated in Fig 3. Integral image makes calculating a feature extremely fast and efficient. It also means calculating features of various sizes requires the same effort as a feature of only two or three pixels. The detection of various sizes of the same object requires the same amount of effort and time as objects of similar sizes since scaling requires no additional effort.

C. Training Classifiers for Facial Features

HAAR Classifier needs to train to detect human facial features, Such as the mouth, eyes, and nose. To train the classifiers, AdaBoost algorithm and HAAR feature algorithms must be implemented. Intel developed an open source library devoted to easing the implementation of computer vision related programs called Open Computer Vision Library (OpenCV). To train the classifiers, two set of images are needed that are negative image set and positive image set. Negative image set contains an image or scene that does not contain the object, in our case a facial feature, which is going to be detected. The other set of images, the positive images, contain objects that are facial features. The location of the objects within the positive images is specified by: image name, the upper left pixel and the height, and width of the object. For training facial features 5,000 negative images with at least a mega-pixel resolution were used. These images consisted of everyday objects, like paperclips, and of natural scenery, like photographs of forests and mountains but not the face image. For more accurate facial feature detection, the original positive set of images is needed which include large variation between different people, including, race, gender, and age. Three separate classifiers were trained, one for the eyes, one for the nose, and one for the mouth. Once the classifiers were trained, they were used to detect the facial features within another set of images in database. The accuracy of the classifier was as shown in Table II.

TABLE II
ACCURACY OF CLASSIFIERS

Facial Feature	**Positive Hit Rate**	**Negative Hit Rate**
Eyes	93%	23%
Nose	100%	29%
Mouth	67%	28%

D. Regionalized Detection

A method is needed to reduce the false positive rate of the classifier and to increase accuracy, without modifying the classifier training attribute. The proposed method is to limit the region of the image that is analysed for the facial features. For example region analysis for mouth is limited to bottom area of face, for nose it is limited to centre area of face and for eyes it is limited to upper area of face. By reducing the area analysed, accuracy will increase since less area exists to produce false positives. It also increases efficiency since fewer features need to be computed and the area of the integral images is smaller.

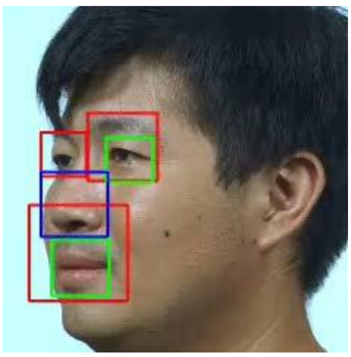

Fig. 4 Inaccurate Detection: Eyes (red), Nose (blue), and Mouth (green),

In order to regionalize the image, one must first determine the likely area where a facial feature might exist. The simplest method is to perform facial detection on the image first. The area containing the face will also contain facial features. However, the facial feature cascades often detect other facial features as illustrated in Fig 4. To eliminate inaccurate feature detection, it can be assumed that the eyes will be located near the top of the head, the nose will be located in the centre area and the mouth will be located near the bottom. The upper 5/8 of the face is analysed for the eyes. The centre of the face, an area that is 5/8 by 5/8 of the face, was used to for detection of the nose. The lower half of the facial image was used to detect the mouth.

After Regionalization the accurate results for feature detection are shown in Fig. 5. Since each the portion of the image used to detect a feature become smaller than that of the whole image, detection of all three facial features takes less time on average than detecting the face itself. Regionalization provides a tremendous increase in efficiency in facial feature detection also increases the accuracy of the detection. All false positives were eliminated. Detection rate is around 95% for nose and eye. The mouth detection has a lower rate due to the minimum area for detection. By changing the height and width parameter to more accurately represent the dimensions of the mouth and retraining the classifier the accuracy should increase the accuracy to that of the other features.

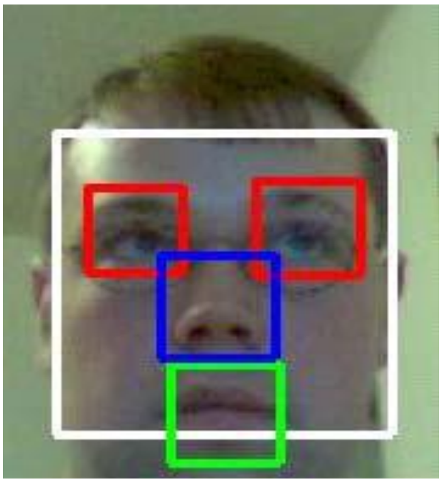

Fig. 5 Detected Objects: Face (white), Eyes (red), Nose (blue) and Mouth (green),

IV. DATA

A. *Introduction*

There are many face datasets available for face recognitions and many of them contain labelled celebrities. Some of them are accessible for all for research purpose, CFW is one of them. Some of the datasets are compared here.

The CFW dataset is advantageous in several aspects [7].

- First, it contain more images of celebrities which much larger than previous datasets, most of which contain only tens of thousands of faces.
- Second, since all faces in CFW are collected from the web, they have large variation in pose, expression, hairstyle and makeup.
- Finally, the labels of CFW dataset are much more accurate than previous automatically generated datasets. For example, the error rate of the labels of Faces in the News dataset is 23%, while the overall error rate of the labels of CFW dataset is 13.93% and a significant portion of CFW dataset (constituting over half of the CFW dataset) achieves an error rate as low as 4.07%.

TABLE III
SCALE COMPARISON WITH EXISTING FACE DATASETS

Database	**#People**	**#Image**
CFW	421,436	2,453,402
Caltech Web Faces	≈10000	10000
CAS-PEAL	1040	99594
FRGC	< 688	50000
FERET	1199	14126
PIE,CMU	68	41368
LFW	5749	13233

Properties of CFW Database are as follows:

- Scale: The CFW dataset contains 2.45 million distinct images of 421436 individuals. Here we compare the scale of CFW database with popular large-scale face datasets in Table III. From the table, it can be observed that the CFW database is much larger than previous face databases in terms of both individuals and images.

- Accuracy: In Table IV, we compare the accuracy of CFW with two other automatically labelled face datasets Faces in the News and Labelled Yahoo News.

We can see that the scale and accuracy of CFW dataset is much better than other datasets. Besides, we can see that a significant portion of the CFW database is comprised of SFSN (Single Face Single Name) images with 95.93% label accuracy, which is far ahead of the results in previous work. Moreover, we can apply a variety of thresholds on the annotation confidences of SFSN faces to get error rates as low as 1.69% while still including a large number of faces compared to other automatically generated dataset, as shown in Table V.

Finally, Table VI lists the top five celebrities with the largest numbers of distinct images in CFW database.

TABLE IV
ACCURACY COMPARISON WITH OTHER AUTOMATICALLY LABELLED FACE DATASETS

Dataset	#Image	Error Rate
CFW-SFSN	1,429,878	4.07%
CFW	2,453,402	13.93%
Faces in News	30,281	23.00%
Labelled Yahoo News	28,204	20.58%

TABLE V
MOST CONFIDENT SFSN IMAGES GIVE SMALL ERROR RATE

#SFSN Faces	Error Rate
1,429,878	4.07%
857,927	2.71%
571,951	2.23%
428,963	1.95%
214,481	1.69%

TABLE VI
TOP FIVE CELEBRITIES WITH THE LARGEST NUMBER OF DISTINCT IMAGES IN CFW

Name	#Distinct Image	Occupation
Britney Spears	11,379	Singer
Barack Obama	11,034	Politician
Angelina Jolie	10,983	Actress
Harry Potter	6,952	Film Character
Lindsay Lohan	6,502	Actress

B. Dataset Descriptions

MSRA-CFW (Microsoft Research Asia- Celebrities on the Web) [13] is a data set of celebrity face images collected from the web. Starting from any face image, its near-duplicate images and associated surrounding texts is obtained. Then the dominant people names by matching with a large list of celebrity names from public websites such as Wikipedia are detected. A classifier is applied to further identify the celebrities appearing in the web images. The dataset includes image URLs faces. To facilitate downloading the images, a number of URLs for the near-duplicates of each face are provided. Besides, the thumbnail images and facial features are also provided for visualization and benchmarking purposes. In the dataset, the files for each person are put into the same folder under this person's name. The files in each folder are categorized into four types:

- Thumbnails: Down sampled images for the faces. These are for visualization purposes only. Please note that each image contains only one face as detected by a face detector
- Info.txt: Contains the "original web images" (OWI) for the thumbnail images. For each thumbnail, info.txt contains a line of metadata followed by a list of near-duplicate image URLs. The metadata consists of: the number of near-duplicate image URLs, the file name of the corresponding thumbnail and the URL of the OWI.
- Feature.bin: contains LBP (Local Binary Pattern) features for the faces. The file starts with two int32 variables indicating the total number of faces and the dimension of LBP features, followed by a byte buffer storing all the features (one face after another).

- Filelist_LBP.txt: each line of this file contains a file name, corresponding to the order of the features in feature.bin. The other four numbers on each line are the location of the faces in the OWIs, where the thumbnails are down sampled (left, top right, bottom).

V. CONCLUSION

This paper introduces different approach for face detection which can be combined for better results. HAAR is feature based method for face detection. HAAR features, Integral images, regionalized detection of features improve Face detection in terms of speed and accuracy. HAAR algorithm also gives small false positives rate. We have also seen CFW dataset which contain large number of celebrity images. CFW is open to all for research purpose and it is downloadable [13]. It gives better results for name annotation, as it contains celebrities' images in various pose, makeup, and hairstyle.

REFERENCES

[1] T. L. Berg, A. C. Berg, J. Edwards, and D. A. Forsyth, "Who's in the picture," presented at the *NIPS, Vancouver, British Columbia*, 2004.

[2] D. Ozkan and P. Duygulu, "A graph based approach for naming faces in news photos," in *Proc. CVPR*, 2006, pp. 1477–1482.

[3] M. Guillaumin, T. Mensink, J. J. Verbeek, and C. Schmid, "Automatic face naming with caption-based supervision," in *Proc. CVPR*, 2008, pp. 1–8.

[4] P. T. Pham, M.-F. Moens, and T. Tuytelaars, "Cross- media alignment of names and faces," *IEEE Trans. Multimedia*, vol. 12, no. 1, pp. 13–27, Jan. 2010.

[5] Ion Marqu´es "Face Recognition Algorithms" June 16, 2010

[6] Phillip Ian Wilson Dr. John Fernandez "Facial Feature Detection Using HAAR Classifiers" *Jcsc* 21, 4 (April 2006)

[7] Xiao Zhang, Lei Zhang, Xin-Jing Wang, Heung-Yeung Shum "Finding Celebritie in Billions of Web Images" *IEEE Transactions On Multimedia*, Vol. 14, No. 4, August 2012

[8] R. Padilla, C. F. F. Costa Filho and M. G. F. Costa "Evaluation of HAAR Cascade Classifiers Designed for Face Detection" *World Academy of Science, Engineering and Technology* 64 2012.

[9] P. Tirilly, V. Claveau, and P. Gros, "News image annotation on a large parallel text- image corpus," presented at the *LREC*, Malta, 2010

[10] Lei Zhang, Longbin Chen, Mingjing Li, H. Zhang "Bayesian Face Annotation in Family Albums" *Microsoft Research Asia*

[11] Xin-Jing Wang, Lei Zhang, Ming Liu, Yi Li, Wei-Ying Ma "ARISTA - Image Search to Annotation on Billions of Web Photos" *Microsoft Research Asia*

[12] Jitao Sang, Changsheng Xu "Robust Face-Name Graph Matching for Movie Character Identification" *IEEE Transactions On Multimedia*, 2010

[13] http://research.microsoft.com/en-us/projects/msra-cfw/default.aspx

