ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

Impact of Violajones Algorithm with Cascade Object Detection on Face Annotation

S.M. Ali[1], Bushra Firdous[2]

Associate Professor, Dept. of ECE, Anjuman College of Engineering and Technology, Nagpur, India[1]

M. Tech Scholar, Dept. of ECE, Anjuman College of Engineering and Technology, Nagpur, India[2]

ABSTRACT: This method can be observed as a type of multi-class image classification with a very huge no. of classes - as large as the vocabulary amount. Generally, image inquiry in the form of extracted feature vectors and the coaching annotation words are used by machine learning approach to attempt to automatically handle annotations to newly added images. The first methods learned the interaction between image features and training annotations, then techniques were developed using machine adaptation to try to translate the textual vocabulary with the 'visual vocabulary', or clustered regions known as blobs. Work following these efforts has included classification approaches, relevance models and so on.
The advantages of automatic image annotation versus content-based image retrieval (CBIR) are that queries can be more naturally specified by the user. CBIR generally (at present) requires users to search by image concepts such as color and texture, or finding example queries. Certain image features in example images may override the concept that the user is really focusing on. The traditional methods of image retrieval such as those used by libraries have relied on manually annotated images, which is expensive and time-consuming, especially given the large and constantly growing image databases in existence.

KEYWORDS: Face annotation, violajones algorithm, machine learning, label refinement, cascade object detection

I.INTRODUCTION

Due to the idolization of diverse digital cameras and the increasing growth of social media tools for internet-based photographs sharing, recent years have witnessed an explosion of the number of digital photos clicked and saved by consumers. A large portion of photographs shared by users on the Internet are human facial images. Some of these facial images are tagged with names, but many of them do not tagged properly [1].

This has inspired the inspection of auto face annotation, an important approach that intent to annotate facial images automatically. For example, with auto face annotation techniques, online photograph-sharing sites (e.g., Face book) can automatically annotate user’s uploaded photographs to expedite online photograph search and management. Besides, face annotation can also be adapted in news video territory to catch important persons appeared in the videos to expedite news video retrieval and summarization tasks[5].

Classical face annotation approaches are often conducted as an lengthy face recognition problem, where various classification models are trained from a collection of bag of features of facial images by employing the supervised or semi-supervised machine learning techniques. However, the “model-based face annotation” techniques are limited in several aspects [2].

First, it is usually time-consuming and costly to collect a large amount of human-labeled training facial photographs. Second, it is usually hard to generalize the models when new data or new persons are added, in which an intensive retraining process is usually compulsory. Last but not least, the annotation performance often scales poorly when the number of persons is very large [6].

Recently, some rising studies have attempted to analyze a promising search-based annotation paradigm for facial image annotation by scooping the World Wide Web (WWW), where a enormous number of bag of features of facial images

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

are freely available. Instead of training explicit classification models by the regular model-based face annotation approaches, the search-based face annotation (SBFA) paradigm aims to tackle the automated face annotation task by exploiting content-based image retrieval (CBIR) techniques in scooping enormous bag of features of facial images on the web. The SBFA framework is data-driven and model-free, which to some extent is inspired by the search-based image annotation techniques for generic image annotations. The main objective of SBFA is to assign correct name labels to a given query facial image. In particular, given a novel facial photo for annotation, we first retrieve a short list of top K-most similar facial images from a bag of features of facial image database, and then annotate the facial image by performing voting on the labels associated with the top K similar facial images.

In particular, we propose a novel unsupervised label refinement (URL) scheme by exploring machine learning techniques to enhance the labels purely from the weakly labeled data without human manual efforts. We also propose a clustering-based approximation (CBA) algorithm to improve the efficiency and scalability. As a summary, the main contributions of this paper include the following:

- We investigate and implement a promising search based face annotation scheme by mining large amount of weakly labeled facial images freely available on the WWW.
- We propose a novel ULR scheme for enhancing label quality via a graph-based and low-rank learning approach.
- We propose an efficient clustering-based approximation algorithm for large-scale label refinement problem.
- We conducted an extensive set of experiments, in which encouraging results were obtained.

We note that a short version of this work had appeared in SIGIR2011. This journal article has been significantly extended by including a substantial amount of new content. The remainder of this paper is organized as follows: Section 2 reviews the related work. Section 3 gives an overview of the proposed search-based face annotation framework. Section 4 presents the proposed unsupervised label refinement scheme.

II. PROPOSED WORK

Our work is related to various groups of research work.

The first group of related work is on the topics of face recognition and authentication, which are classic analysis complication in computer eyesight and pattern recognition and have been extensively studied for many years .Recent years have observed some emerging criterion studies of unconstrained face detection and verification techniques on facial photographs that are merge from the web, such as the LFW criterion studies. Some recent study had also attempted to increase classic face recognition techniques for face annotation tasks [8]. Complete reviews on face recognition and authentication topics can be found in some survey papers and books.

The second group is about the studies of generic image annotation. The classic image annotation approaches usually apply some existing object recognition techniques to train allocation models from human-features training images or attempt to assume the correlation/probabilities between images and annotated keywords. Given finite training data, semi-supervised learning methods have also been used for image annotation. For example, Wang et al. proposed to refine the model-based annotation results with a label similarity graph by following random walk principle. Similarly, Pham et al. proposed to annotate unlabeled facial images in video frames with an iterative label propagation scheme. Although semi-supervised learning approaches could leverage both labeled and unlabeled data, it remains fairly time-consuming and expensive to collect enough well-labeled training data to achieve good performance in large-scale scenarios. Recently, the search-based image annotation paradigm has attracted more and more attention [7]. For example, Russell et alia. Built a huge assemblage of web images with ground truth features to expedite object acceptance research. However, most of these works were concentrating on the ratio, search, and bag of feature techniques. dissimilar these existing works, we propose a unsupervised features clarification scheme that is concentrate on optimizing the features quality of facial images towards the content-based face annotation task.

The third group is about face annotation on personal/family/social photos. Several studies have mainly focused on the annotation task on personal photos, which often contain rich contextual clues, such as personal/family names, social context, geotags, and time stamps and so on. The number of persons/classes is usually quite small, making such

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

annotation tasks less challenging. These techniques usually achieve fairly accurate annotation results, in which some techniques have been successfully deployed in commercial applications [9], for example, Apple photos, Google Picasa, Microsoft easy Album, and Face book face auto tagging solution.

The fourth group is about the studies of face annotation in mining weakly labeled facial images on the web. Some studies consider a human name as the input query, and mainly aim to refine the text-based search results by exploiting visual consistency of facial images. For example, Oscan and Duygul proposed a graph-based model for finding the densest sub-graph as the most related result. Following the graph-based approach, Le and Satoh proposed a new local density score to represent the importance of each returned images, and Guillemin et al. introduced a modification to incorporate the constraint that a face is only depicted once in an image [4]. On the other hand, the generative approach like the Gaussian mixture model was also been adopted to the name-based search scheme and achieved comparable results. Recently, a discriminate approach was proposed in to improve over the generative approach and avoid the explicit computation in graph-based.

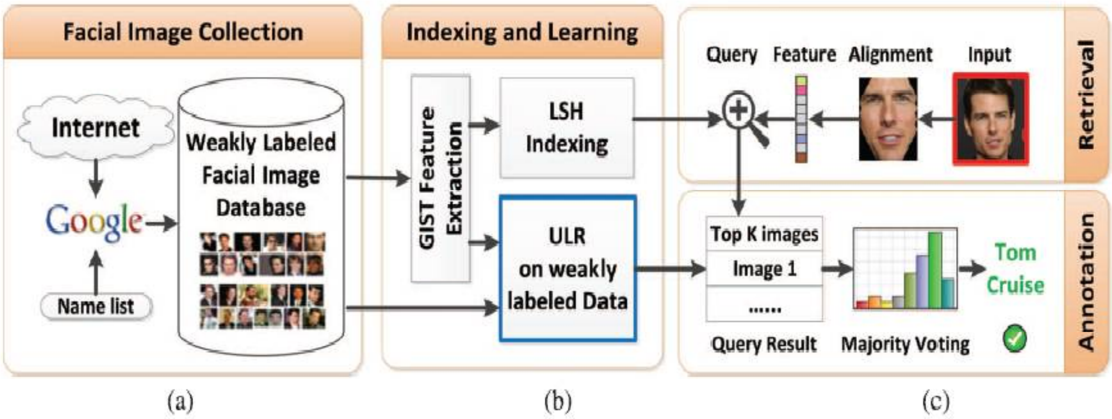

Fig. 1. The system flow of the proposed search-based face annotation scheme. (a) We collect weakly labeled facial images from WWW using web search engines. (b) We preprocess the crawled web facial images, including face detection, face alignment, and feature extraction for the detected faces; after that, we apply LSH to index the extracted high-dimensional facial features. We apply the proposed ULR method to refine the raw weak labels together with the proposed clustering-based approximation algorithms for improving the scalability. (c) We search for the query facial image to retrieve the top K similar images and use their associated names for voting toward auto annotation.

By using ideas from query expansion, the performance of name-based scheme can be further improved with introducing the images of the “friends” of the query name. Unlike these studies of filtering the text-based retrieval results, some studies have attempted to directly annotate each facial image with the names extracted from its caption information. For example, Berg et al. proposed a possibility model combined with a clustering algorithm to estimate the relationship between the facial images and the names in their captions. For the facial images and the detected names in the same document (a web image and its corresponding caption), Guillemin et al. proposed to iteratively update the assignment based on a minimum cost matching algorithm. In their follow-up work, they further improve the annotation performance by using distance metric learning techniques to achieve more discriminative feature in low-dimension space. Our work is different from the above previous works in two main aspects. First of all, our work aims to solve the general content-based face annotation problem using the search-based paradigm, where facial images are directly used as query images and the task is to return the corresponding names of the query images.
Very limited research progress has been reported on this topic. Some recent work mainly addressed the face retrieval problem, in which an effective image representation has been proposed using both local and global features. Second, based on initial weak labels, the proposed unsupervised label refinement algorithm learns an enhanced new label matrix for all the facial images in the whole name space; however, the caption-based annotation scheme only considers the assignment between the facial images and the names appeared in their corresponding surrounding-text. As a result, the caption-based annotation scheme is only applicable to the scenario where both images and their captions are available, and cannot be applied to our SBFA framework due to the lack of complete caption information.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

III. BLOCK DIAGRAM

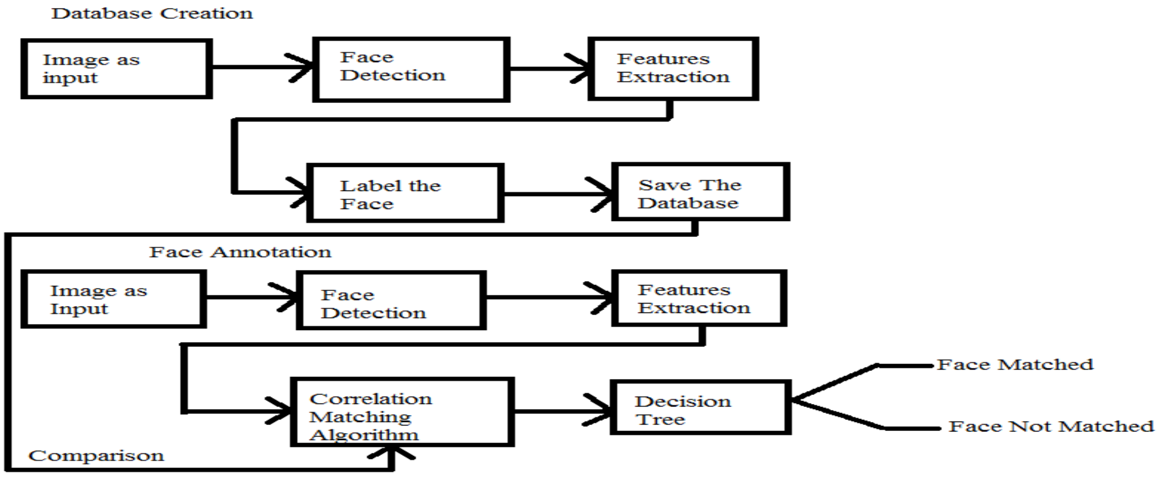

Figure 3.1.1 Block diagram for database creation and Face Annotation

A. Database Creation

Firstly, for achieving our goal of face annotation we require a database for that we use the data set of Delhi because it is having more Indian faces, so for creating the database we have to pick an image from data set, let this image be an input for our system, after picking an image we detect the face and mark the bounding box denoted by the black boxes in figure 3.1.2, simultaneously extract the features of the face like Eye width, Eye height, Nose width, Nose height, Mouth width, Mouth height, Eye to Eye distance, Left Eye to Nose distance, Right Eye to Nose distance, Left Eye to Mouth distance, Right Eye to Mouth distance denoted by red, yellow, blue and green lines in the figure 3.1.2, this all we have done by violajones algorithm with cascade object detection. After that we have to label the image for our database or for our convenience. Here, now we have created the database and now ready to proceed further.

Figure 3.1.2 Database Creation

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

B. Face Annotation

Now it's time to annotate the face which we have saved in the database or which we want to annotate. The first 3 stages are same as we have done in creating the database that is we want an image as an input to find out whether we have that image in database or not 2nd and 3rd stage that is face detection and features extraction by violajones algorithm with cascade object detection is also same as shown in the figure 3.1.2.

The most important part in this bock diagram is comparison of our image and the image saved in database. This comparison will be done by Correlation Matching algorithm, which compares the features of an image with the image that saved in data base like, Eye width, Eye height, Nose width, Nose height, Mouth width, Mouth height, Eye to Eye distance, Left Eye to Nose distance, Right Eye to Nose distance, Left Eye to Mouth distance, Right Eye to Mouth distance denoted by red, yellow, blue and green lines in the figure 3.1.3, and it gives these values to decision tree which decides whether the image has been matched with database or not and give us the result as shown in figure 3.1.3.

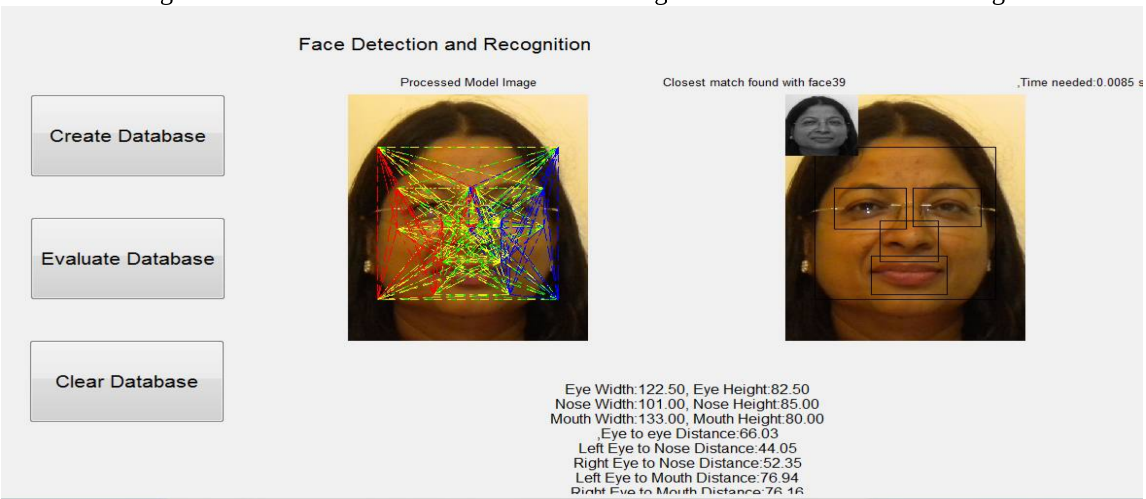

Figure 3.1.3 Face Annotation or the result

IV. VIOLAJONES ALGORITHM WITH CASCADE OBJECT DETECTION

The Viola–Jones object detection groundwork is the first object detection groundwork to provide competing object detection rates in real-time, it was inspire primarily by the complication of face detection. The problem to be solved is detection of faces in an image. A human can do this easily, but a computer needs precise instructions and constraints. To make the task more manageable, Viola–Jones requires full view frontal upright faces. Thus in order to be detected, the entire face must point towards the camera and should not be tilted to either side. While it seems these constraints could diminish the algorithm's utility somewhat, because the detection step is most often followed by a recognition step, in practice these limits on pose are quite acceptable [10].

The characteristics of Viola–Jones algorithm which make it a good detection algorithm are:

- Robust – very high detection rate (true-positive rate) & very low false-positive rate always.Real time – For practical applications at least 2 frames per second must be processed.
- Face detection only (not recognition) - The goal is to distinguish faces from non-faces (detection is the first step in the recognition process).

The algorithm has four stages: i) Haar Features – All human faces share some similar properties. These regularities may be matched using Haar Features.

A few properties common to human faces:

- The eye region is darker than the upper-cheeks.
- The nose bridge region is brighter than the eyes.

Composition of properties forming match able facial features:

- Location and size: eyes, mouth, bridge of nose
- Value: oriented gradients of pixel intensities

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

The four features matched by this algorithm are then sought in the image of a face[10].
Rectangle features:

- Value = Σ (pixels in black area) - Σ (pixels in white area)
- Three types: two-, three-, four-rectangles, Viola & Jones used two-rectangle features
- For example: the difference in brightness between the white &black rectangles over a specific area
- Each feature is related to a special location in the sub-window

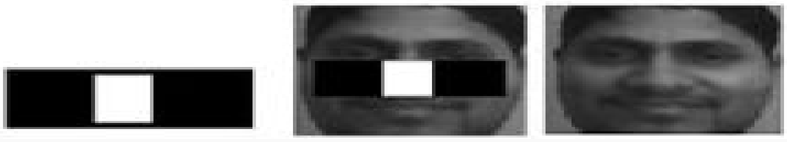

Fig-4.2.1.a: Haar Feature that looks similar to the bridge of the nose is applied onto the face

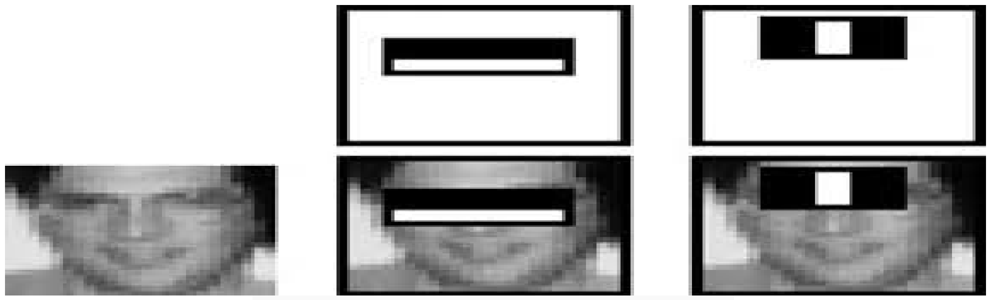

Fig-4.2.1.b: 3rd and 4th kind of Haar Feature

ii) Creating an Integral Image – An image representation called the integral image evaluates rectangular features in constant time, which gives them a considerable speed advantage over more sophisticated alternative features. Because each feature's rectangular area is always adjacent to at least one other rectangle, it follows that any two-rectangle feature can be computed in six array references, any three-rectangle feature in eight, and any four-rectangle feature in nine[1][2].
iii) Adaboost Training – The speed with which features may be evaluated does not adequately compensate for their number, however. For example, in a standard 24x24 pixel sub-window, there are a total of M=162,336 possible features, and it would be prohibitively expensive to evaluate them all when testing an image. Thus, the object detection framework employs a variant of the learning algorithm AdaBoost to both select the best features and to train classifiers that use them. This algorithm constructs a "strong" classifier as a linear combination of weighted simple "weak" classifiers.
iv) Cascading Classifiers – In cascading, each stage consists of a strong classifier. So all the features are grouped into several stages where each stage has certain number of features.
The job of each stage is to determine whether a given sub-window is definitely not a face or may be a face. A given sub-window is immediately discarded as not a face if it fails in any of the stages [10].
A simple framework for cascade training is given below:

- User selects values for f, the maximum acceptable false positive rate per layer and d, the minimum acceptable detection rate per layer.
- User selects target overall false positive rate Ftarget.
- P = set of positive examples
- N = set of negative examples

➢ To Detect a Feature:

1. Define and set up your cascade object detector using the constructor.
2. Call the step method with the input image, I, the cascade object detector object, detector, points PTS, and any optional properties. See the syntax below for using the step method.

Use the step syntax with input image, I, the selected Cascade object detector object, and any optional properties to perform detection.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

BBOX = step (detector, I) returns BBOX, an M-by-4 matrix defining M bounding boxes containing the detected objects. This method performs multiscale object detection on the input image, I. Each row of the output matrix, BBOX, contains a four-element vector, [x y width height], that specifies in pixels, the upper-left corner and size of a bounding box. The input image I, must be a grayscale or true color (RGB) image.

BBOX = step (detector, I, roi) detects objects within the rectangular search region specified by roi. You must specify roi as a 4-element vector, [x y width height], that defines a rectangular region of interest within image I. Set the 'Use ROI' property to true to use this syntax[10].

V. RESULT AND DISCUSSION

In the table 1, it shows the table of number of entries in database, mean delay, number of correct evaluation and accuracy to analysis the result.

No. of entries in database	Mean delay	No. of Correct Evaluation	Accuracy
10	0.0209	10	100%
20	0.0613	20	100%
30	0.1177	30	96.60%
40	1.0353	40	100%
50	1.1465	50	100%
60	1.2639	60	100%
70	1.3969	70	100%

Table.1 Result analysis

In the fig.2 shows number of entries in database vs mean delay.

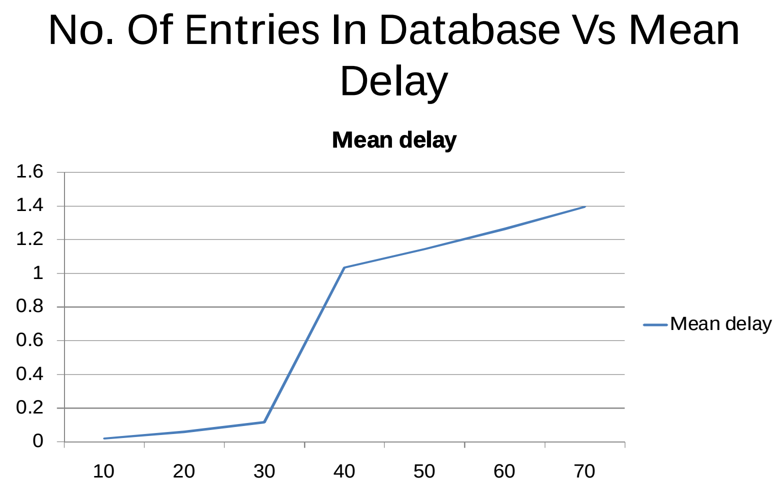

fig.2 No. Of entries in database vs mean dealy

In the fig.3 shows number of correct evaluation vs accuracy by this graph we get our result.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

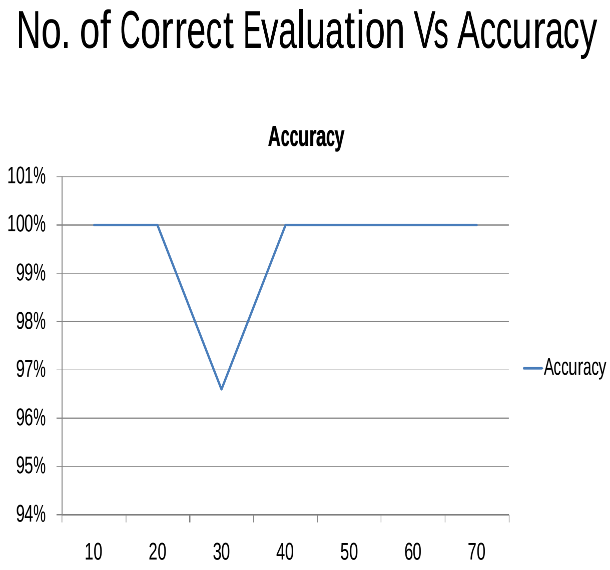

Fig.3 no. Of correct evaluation vs accuracy

VI.CONCLUSION

We investigated a promising search-based face annotation framework, in which we focused on tackling the critical problem of enhancing the label quality and proposed a violajones algorithm with cascade object detection. From an extensive set of experiments, we found that the proposed technique achieved promising results under a variety of settings. Our experimental results also indicated that the proposed violajones algorithm with cascade object detection significantly surpassed the other regular approaches in literature and give more accuracy compare to others in literature.

REFERENCES

[1] Mining Weakly Labeled Web Facial Images for Search-Based Face Annotation Dayong Wang, Steven C.H. Hoi, Member, IEEE, Ying He, and Jianke Zhu, January 2014.

[2]Enhancing performance of face recognisation system by using near set approach for selecting facial features Shivani Gupta, et. al., 2005-2008]

[3] Social Media Modeling and Computing, S.C.H. Hoi, J. Luo, S. Boll, D. Xu, and R. Jin, eds. Springer, 2011.

[4] L. Zhang, L. Chen, M. Li, and H. Zhang, "Automated Annotation of Human Faces in Family Albums," Proc. 11th ACM Int'l Conf. Multimedia (Multimedia), 2003.

[5] T.L. Berg, A.C. Berg, J. Edwards, M. Maire, R. White, Y.W. Teh, E.G. Learned-Miller, and D.A. Forsyth, "Names and Faces in the News," Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR), pp. 848-854, 2004.

[6] J. Yang and A.G. Hauptmann, "Naming Every Individual in News Video Monologues," Proc. 12th Ann. ACM Int'l Conf. Multimedia (Multimedia), pp. 580-587. 2004.

[7] J. Zhu, S.C.H. Hoi, and M.R. Lyu, "Face Annotation Using Transductive Kernel Fisher Discriminant," IEEE Trans. Multimedia, vol. 10, no. 1, pp. 86-96, Jan. 2008.

[8] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-Based Image Retrieval at the End of the Early Years," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1349-1380, Dec. 2000.

[9] Face recognition: a convolution neural network approach [NEC Research Institute, Princeton, et. al., 1997]

[10] Website:
1.en.wikipedia.org/wiki/Viola%E2%80%93Jones_object_detection_framework
2.in.mathworks.com/help/vision/ref/vision.cascadeobjectdetector.