5-2013

FANS: Face Annotation by Searching Large-scale Web Facial Images

Steven HOI
Singapore Management University, CHHOI@smu.edu.sg

Dayong WANG
Nanyang Technological University

I Yeu CHENG
Nanyang Technological University

Elmer LIN
Nanyang Technological University

Jianke ZHU
Zhejiang University

See next page for additional authors

DOI: https://doi.org/10.1145/2487788.2487933

FANS: Face Annotation by Searching Large-scale Web Facial Images

Steven C.H. Hoi, Dayong Wang,
I Yeu Cheng, Elmer Weijie Lin, Jianke Zhu‡, Ying He, Chunyan Miao
School of Computer Engineering, Nanyang Technological University, Singapore
‡College of Computer Science, Zhejiang University, Hangzhou, 310027, P.R. China
{chhoi, s090023, p090122, linw0028, yhe, ascymiao}@ntu.edu.sg, jkzhu@zju.edu.cn

ABSTRACT

Auto face annotation is an important technique for many real-world applications, such as online photo album management, new video summarization, and so on. It aims to automatically detect human faces from a photo image and further name the faces with the corresponding human names. Recently, mining web facial images on the internet has emerged as a promising paradigm towards auto face annotation. In this paper, we present a demonstration system of search-based face annotation: **FANS** — Face ANnotation by Searching large-scale web facial images. Given a query facial image for annotation, we first retrieve a short list of the most similar facial images from a web facial image database, and then annotate the query facial image by mining the top-ranking facial images and their corresponding labels with sparse representation techniques. Our demo system was built upon a large-scale real-world web facial image database with a total of $6,025$ persons and about 1 million facial images. This paper demonstrates the potential of searching and mining web-scale weakly labeled facial images on the internet to tackle the challenging face annotation problem, and addresses some open problems for future exploration by researchers in web community. The live demo of FANS is available online at `http://msm.cais.ntu.edu.sg/FANS/`.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; I.2.6 [**Artificial Intelligence**]: Learning

General Terms

Algorithms, Experimentation

Keywords

web facial images, search-based face annotation, web data mining

1. INTRODUCTION

With the rapid growth of web photo sharing portals and social networks, massive amounts of images and photos have been uploaded and shared on the internet nowadays. A considerable amount of these online photos is related to facial images as they are closely related to social activities of human beings. The huge amount of weakly labeled or unlabeled facial images poses many emerging challenges and research opportunities. This paper addresses one of the open research problems, "auto face annotation", i.e., how to automatically assign a face with the name of the corresponding person. This technique is beneficial to many real-world applications, including online photo album management, new video summarization, and so on.

Typically, face annotation can be solved by applying classical face recognition methods, which have been extensively studied. In particular, one can apply supervised machine learning techniques to train face classification models from a collection of well-controlled labeled facial images and apply the model to name a new facial image. In recent years, face recognition and annotation techniques have been widely used in commercial personal/family photo management, e.g., Apple iPhoto,[1] Google Picasa,[2] and Facebook face auto-tagging solution.[3] Although conventional face annotation algorithms have been extensively studied, they all suffer from some common drawbacks, e.g., the requirement of high-quality training data and poor scalability for handling large number of persons.

Recent years are witnessing a promising direction to tackle the web-scale face annotation problem. In particular, some studies [6, 9, 1, 3] have attempted to tackle the face annotation challenge by mining vast amounts of weakly labeled facial images freely available on the internet. Unlike the regular face annotation approaches, such data-driven annotation techniques exploit the massive amount of weakly labeled data in annotating the facial images directly without building explicit classification models, which is able to overcome some limitations of classical face annotation techniques.

In this paper, we focus our attention on the search-based face annotation paradigm. Specifically, given a user-uploaded facial image for annotation, the search-based face annotation scheme firstly retrieves a short list of top-K most similar facial images from a large-scale web facial image database, and then annotates the query facial image by mining the labels associated with the top-K similar facial images. In general, the search-based face annotation scheme has to tackle two main challenges [6, 9, 8, 10]: (i) How to efficiently retrieve the top-K most similar facial images from a large facial image database given a query facial image, i.e., how

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

[1] `http://www.apple.com/ilife/iphoto/`
[2] `http://picasa.google.com/`
[3] `http://www.facebook.com/`

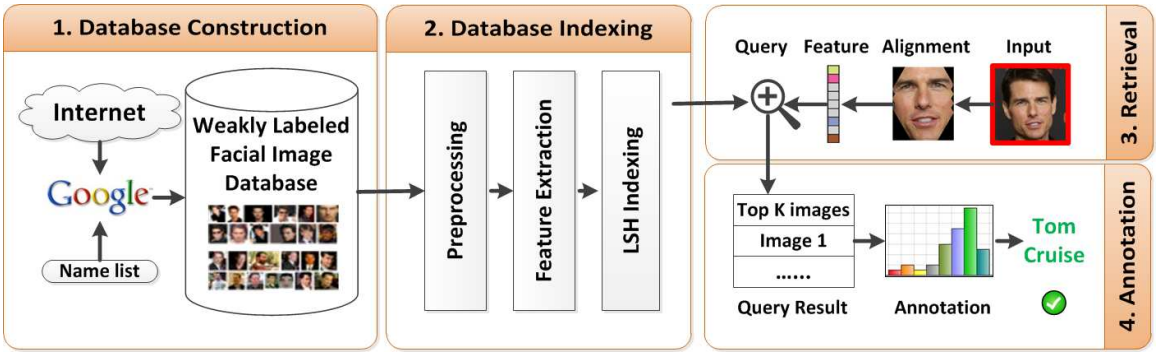

Figure 1: The Framework of the FANS system.(1) The database construction by crawling facial images from the World Wide Web; (2) The database indexing for fast facial feature retrieval in high-dimensional space; (3) The content-based facial image retrieval for a query facial image; (4) The automatic face annotation by mining the retrievable facial images and the corresponding labels.

to develop an effective content-based facial image retrieval solution; and (ii) How to effectively exploit the short list of candidate facial images and their weak labels for naming the faces automatically.

Following our recent work [6, 9, 8], in this demo paper, we present the implementation of an automated search-based face annotation system: **FANS** — Face ANnotation by Searching large-scale web facial images. In particular, the **FANS** system is built on a real-world large-scale web facial images database, which contains 6,025 famous western celebrities and about 1-million facial images. It adopts the Locality-Sensitive Hashing (LSH) [2] for high-dimensional indexing to facilitate the task of similar face retrieval, and exploits the state-of-the-art sparse representation techniques for face annotation. The rest of this paper is organized as follows. Section 2 presents the framework of the **FANS** system and the techniques for each module in detail. Section 3 shows the user-interfaces (UI) of FANS. Section 4 discusses the evaluations and applications, and Finally Section 5concludes this paper.

2. FRAMEWORK OF FANS SYSTEM

In this section, we briefly introduce the framework of our FANS system. Figure 1 illustrates the proposed framework that consists of the following four major modules: (1) the database construction module by crawling facial images from the World Wide Web; (2) the database indexing module towards fast retrieval of high-dimensional facial features; (3) the content-based facial image retrieval module for searching a query facial image; (4) the automated face annotation module for naming the query by mining the top-k retrieved similar facial images and their corresponding weakly labels. The details of each module are described as follows.

The first module, as shown in Figure 1 (1), is to collect a large database of web facial images which are freely available on the World Wild Web. Although some web facial images database are available in the internet, e.g., LFW,[4] [4] Pubfig,[5] [5] and Yahoo! News [3][6]. They are not suitable for the search-based face annotation scheme of our FANS system, as the number of images per person is too small. In our work, we construct our own web facial image database by crawling facial images with web image search engines. Our database consists of 6,025 famous western celebrities and a total of about 1-million facial images. There are two main steps to build such a web facial image database: (i) constructing a name list of popular celebrities; and (ii) querying the existing web search engines with the celebrity names, and then crawl the web facial images according to the web search results. In our approach, we first collect a name list with 6,025 celebrity names downloaded from the website of **IMDb** [7] with the billboard: "Most Popular People Born In yyyy", where yyyy is the born year. Our name list covers the actors and actresses who were born between 1950 and 1990. After that we submitted each name on the lists as a query to search for related web facial images by Google image search engine, and crawled the top 400 returned web images for each query name.

The second module, as shown in Figure 1(2), is to preprocess and index the web facial image database, including face detection, alignment, facial feature representation, and high dimensional feature indexing. We adopt the Viola-Jones algorithm implemented in OpenCV to detect face regions. After that, we aligned all the facial images into a consistent position according to the facial feature points positions by adopting the DLK algorithm [11]. After ignoring the non-face-detected images, we collected about 1 million facial images of 6,025 persons in our web facial image database. For feature representation, we extract the GIST features as the facial feature. According to our experiment, the GIST feature archives closing experimental results with the famous LBP feature with lower feature dimension. Finally, we apply the Locality-Sensitive Hashing (LSH) [2][8] to index the facial features. For the parameter tuning of LSH, we adopt 8 hash tables and 40 bins probed in each hash table. When the recall value is about 0.8, the query time over the 1-million facial image database is around 0.2 second.

The first two modules must be done before a query facial image can be annotated. The next two modules are related to online processes of annotating a query facial image on the fly. As shown in Figure 1(3), given a query facial image, we firstly adopt the same pre-processing step as the one for the aforementioned web facial image database, then we employ a similar face retrieval process to find a short list

[4] http://vis-www.cs.umass.edu/lfw/
[5] http://www.cs.columbia.edu/CAVE/databases/pubfig/
[6] http://lear.inrialpes.fr/people/guillaumin/data.php
[7] http://www.imdb.com
[8] http://lshkit.sourceforge.net/

(a) Query Input Method (b) Retrieval Result Representation

Figure 2: The user interface of the FANS system. (a) The methods of inputting query facial image, including entering the URL, uploading the local image, and directly dragging the images on the webpage; (b) The representation of retrieval result.

of the most similar faces from the indexed face databases using LSH. After obtaining the top-k most similar faces, the last key module is to annotate the query image by mining the short list of top-ranked candidate facial images and their label information, as shown in Figure 1(4). In our approach, we adopt a recently proposed sparse reconstruction algorithm, the weak label regularized local coordinate coding [9], for tackling the automated face annotation task in the FANS system. It can annotate the query facial image by exploiting both weakly label information and visual contents of top-ranked facial images to maximize the annotation efficacy. The details of the proposed annotation and learning techniques can be found in our recent work [9, 8].

3. USER INTERFACES OF FANS

3.1 Query Input Methods

FANS provides three types of query input methods for users to specify a query, which is shown in Figure 2 (a). Similar to most commercial web image search engines, there is an input query box at the top of the front-page of FANS. A user can easily upload a local query image from his/her computer by clicking the camera icon on the right hand side of of the query box. Another input method of FANS is to directly use the web facial images without even downloading them into users' computers. In particular, the second input method is to allow a user specify a query image by simply typing the URL of any facial image on the web in the query box. Sometimes users may want to further explore the retrieval results to pick up some related image for query. Thus, as the third query input method of FANS, users can directly drag any image on the current retrieval page to the input query box for annotation. We allow the user take pictures with their local web-cameras and query their own photos easily.

3.2 Presentation of Retrieval Results

In the retrieval result presentation page of FANS, there are three major parts, including the query image, the annotated name, the corresponding most similar facial image, and the list of top-K similar images. As shown in Figure 2 (b), we present the query image at top-left corner of the result page, and highlight the facial image with red rectangle, under the query image, we show the alignment result of the detected face. If multiple faces are detected in the query image, we will mark all the facial images with color rectangular, and highlight the current query face in red color. At the top-right corner of the result web page, we present the annotated name and its corresponding image. At the bottom of the result web page, we show the top $K = 21$ similar facial images and the their corresponding name info. To browse the details of the thumbnail images of the retrieval results, the users can simply click the thumbnail image to check its original large-size image with higher resolution in a new pop-out window. As a feedback to FANS, the users can help to rate the annotation results by using the rating bar under the similarity value, which can collected the information for further improving the similarity estimation of FANS by adopting machine learning techniques.

With the explosion of social network and social media, more and more personalized photos are shared over the social network, e.g. Facebook, Flicker, Google+ and so on. In our FANS system, the users can login to his/her Facebook account to access his/her own album images and the profile images of his/her friends, then the users can directly drag these images as input queries for finding out the most similar popular star to himself/herself or to his/her friends. The users can also easily share the retrieval results with his/her friends in the Facebook, which is helpful to improve the user interaction for the social network website.

3.3 Mobile Application on Android System

In addition to Web, we deploy FANS on mobile platforms, which has become an important way for people to share and manage their photos. We have developed an App (a.k.a. "Application software") on Android platform, as shown in Fig. 3.

The FANS App has three main functions, as the three buttons shown in Fig. 3. By clicking the first "**SNAP**" button, users can query any photo captured by the cameras of his/her mobile devices. An example of an annotation result is shown in Fig. 3 (b), where the left image is the query image, the right image is the most similar images in the re-

Figure 3: (a) Interface of the mobile application, (b) Annotation result using a captured photo, (c) Edit and crop image, and (d) Access the photo albums on the Facebook.

trieval database. The middle number indicates the similarity value between the query image and the most similar facial image. By clicking the Facebook button below the similar number, users can share the annotation results with their friends via social networks. FANS also allows users to rate the annotation results in FANS App. In particular, under the query image, a bar of five star is given to let users vote the annotation result. For a large size image, users can crop out the facial image region with FANS App as the query image, as shown in Figure 3 (c). By clicking the second button or the third button, users can respectively access the album images in the mobile device or in their Facebook account, so as to explore facial images as queries to interact with FANS, as shown in Figure 3 (d).

4. EVALUATION OF FANS SYSTEM

4.1 Evaluation of Running Time Cost

In FANS, the face detection process is very fast, and the following face alignment process will take about 1.3 seconds per facial image by using the DLK algorithm [11], which depends on the size of the query image. It takes about 0.25 second to extract the facial feature for each facial image. Based on the LSH indexing technique, FANS takes about 0.2 second on average per query for retrieving the top-100 similar images from an one million facial image database. We note that this time cost evaluation is only based on a single regular machine and does not fully explore multi-core computing technique. Finally, the real running time cost of FANS system depends on the network connections and the system load.

4.2 Evaluation of Face Annotation

In order to evaluate face annotation performance of FANS system, we build an evaluation database by randomly choosing 119 names from the whole name list. After that we manually annotate the top-200 to top-400 Google search results and ignore the unrelated images; as a result there are 1,600 facial images in the evaluation dataset. For the retrieval database in the experiments, we use a subset of the whole web facial image database by only collecting the top-200 Google search results from 400 famous persons. As a result, the accuracy of the first annotated name (hit rate @ top-1) is about 60% for the basic majority voting scheme, which can be further improved to about 80% by applying state-of-the-art machine learning techniques [8, 9]. We refer readers to more detailed analysis of annotation performance in our previous work [6, 9, 8, 7].

5. CONCLUSION

This paper presented **FANS**: a novel search-based face annotation system by searching a large-scale web facial images database. By using three kinds of query input methods in the FANS system, one user can easily search for the famous persons and interact with friends in social networks, e.g., Facebook. Besides, we develop an Android APP by deploying our FANS system in the popular smart phone platform. In the future, we plan to improve the accuracy of face retrieval and name annotation by exploring advanced learning techniques.

Acknowledgments

This research is supported in part by MOE tier 1 project grant (RG33/11), and Interactive and Digital Media Programme Office (IDMPO), National Research Foundation (NRF) hosted at Media Development Authority (MDA) under Grant No.: MDA/IDM/2012/8/8-2 VOL 01.

6. REFERENCES

[1] T. L. Berg, E. C. Berg, J. Edwards, and D. A. Forsyth. Who is in the picture. In *NIPS*, pages 264–271. MIT Press, 2006.

[2] W. Dong, Z. Wang, W. Josephson, M. Charikar, and K. Li. Modeling lsh for performance tuning. In *CIKM*, pages 669–678, 2008.

[3] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid. Face recognition from caption-based supervision. *IJCV*, 96(1):64–82, Jan. 2012.

[4] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[5] N. Kumar, A. Berg, P. Belhumeur, and S. Nayar. Attribute and simile classifiers for face verification. In *IEEE ICCV*, pages 365–372, 2009.

[6] D. Wang, S. C. Hoi, and Y. He. Mining weakly labeled web facial images for search-based face annotation. In *ACM SIGIR*, pages 535–544, 2011.

[7] D. Wang, S. C. Hoi, Y. He, and J. Zhu. Mining weakly-labeled web facial images for search-based face annotation. *IEEE TKDE*, 99(PrePrints):1, 2012.

[8] D. Wang, S. C. H. Hoi, and Y. He. A unified learning framework for auto face annotation by mining web facial images. In *CIKM*, pages 1392–1401, 2012.

[9] D. Wang, S. C. H. Hoi, Y. He, and J. Zhu. Retrieval-based face annotation by weak label regularized local coordinate coding. In *ACM Multimedia*, pages 353–362, 2011.

[10] X.-J. Wang, L. Zhang, X. Li, and W.-Y. Ma. Annotating images by mining image search results. *IEEE TPAMI*, 30(11):1919–1932, 2008.

[11] J. Zhu, L. Van Gool, and S. Hoi. Unsupervised face alignment by robust nonrigid mapping. In *IEEE CVPR*, pages 1265–1272, 2009.