IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Conference on Innovative Applications and Research in Computer Science and Engineering (NCIARCSE-2017)

AGTI's Dr. Daulatrao Aher College Engineering, Vidyanagar Extension, Karad

Vol. 4, Special Issue 4, January 2017

Celebrity Face Naming based on Relationships and Knowledge using Caption-Based Supervision

Jyoti H. Jadhav[1], Pankaj Agarkar[2]

Computer Engineering Department, DYPSOE, Pune, India [1,2]

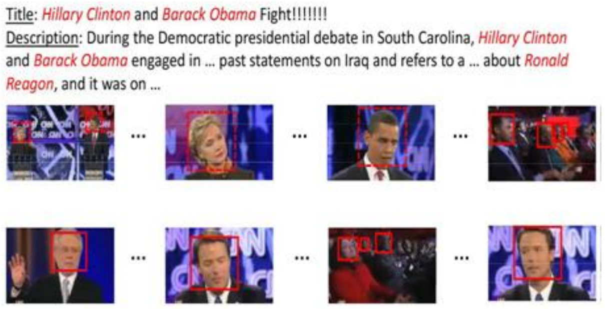
Abstract: Auto face annotation is important in many real world knowledge management system and multimedia information. Face annotation is a field of face detection and recognition. Mining weakly labeled web facial images on the internet has emerged as a promising paradigm towards auto face annotation. The System examines the problem of celebrity face naming in unconstrained video with user provided metadata. Normally we depend on accurate face labels for supervised learning. But sometimes the faces were not properly annotated. One of the solution used in the proposed system is that; it uses the two parameters a rich set of relationships automatically derived from video content and knowledge from image domain. Relationship is the appearance of faces under different context and their visual similarities. The knowledge includes Web images weakly tagged with celebrity names and the celebrity social networks. The relationships and knowledge is elegantly encoded using conditional random field (CRF) for label inference. Two versions of face annotation are considered: within-video and between-video face labeling. The system further rectifies the error in the metadata to correct false labels and annotate the faces with missing names in the metadata of a video by considering a group of socially connected videos for joint label inference. The system leads to higher accuracy in face labeling than several existing approaches but with minor degradation in speed efficiency.

Keywords: Celebrity face naming, face annotation, social network, unconstrained web video, weak label.

I. INTRODUCTION

Digital photo albums and videos are growing explosively in both number and size due to rapid popularization of digital cameras and mobile phone cameras. A large portion of photos and videos are shared by users on Internet. Face annotation for effective management of personal photos in online social networks (OSNs) is currently of considerable practical interest. Labeling celebrities in Web videos is a challenging problem due to large variations in face appearance. These large collections require the annotation of some semantic information to facilitate browsing, manipulation and sharing of photos. The existing OSNs only support manual face annotation, a task that can be considered time-consuming and labor-intensive. The problem becomes increasingly important due to the massive growth of videos in Internet. According to YouTube trends map,1 about 80% of popular videos are people related and among the people-related videos, about 75% are about celebrities. To date, most search engines index these videos with user-provided text descriptions (e.g., title, tag), which are often noisy and incomplete. The descriptions are given globally and hence the correspondences between celebrity names and faces are not explicit. It is not unusual that a mentioned celebrity does not appear in the video, and vice versa, a celebrity actually appearing in a video is not mentioned. For these reasons, searching people-related videos may yield unsatisfactory retrieval performance. The large number of human facial images shared over the different social real world applications some of these images are tagged properly with names, but many of them are not tagged properly. This is motivated the study of auto face annotation. Auto face annotation is an important technique which automatically gives name of relevant person.

Auto face annotation technique is beneficial to many real world applications. For example, with auto face annotation techniques, online photo-sharing sites (e.g. Facebook) can automatically annotate users' uploaded photos to facilitate online search and management. Besides, face annotation can also be applied in news video domain to detect important persons appeared in the videos to facilitate news video retrieval and summarization tasks.



Title: Hillary Clinton and Barack Obama Fight!!!!!!
Description: During the Democratic presidential debate in South Carolina, Hillary Clinton and Barack Obama engaged in ... past statements on Iraq and refers to a ... about Ronald Reagon, and it was on ...

Fig. 1. Example of Web video illustrating the challenge of associating the names in metadata with the detected faces

Fig. 1. illustrates the problem with a real example of Web video. Out of the fourteen faces (of four celebrities) detected in the video, only four of them have names mentioned in the metadata. Furthermore, among the three celebrities who are mentioned, only two of them appear in the video. In other words, there are missing faces and names in the video and text respectively.

DOI 10.17148/IARJSET/NCIARCSE.2017.18

II. RELATED WORK

The first group of related work is on the topics of face recognition and verification, which are classical research problems in computer vision and pattern recognition. Biometric-based technologies include identification based on Physiological characteristics and behavioral traits. Face recognition appears to offer several advantages over other Biometric method; facial images can be easily obtained with a couple of inexpensive fixed cameras. Good face recognition algorithms and appropriate preprocessing of the images can compensate for noise and slight variations in orientation, scale and illumination.

The second group is about the studies of generic image annotation. The classical image annotation approaches usually apply some existing object recognition techniques to train classification models from human-labeled training images or attempt to infer the correlation/probabilities between images and annotated keywords. Given limited training data, semi-supervised learning methods have also been used for image annotation [9], [10], and [11]. For example, Wang et al. [9] proposed to refine the model-based annotation results with a label similarity graph by following random walk principle. Similarly, Pham et al. [10] 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 [3].

The third group is about face annotation on personal/ family/social photos. It focused on the annotation task on personal photos, which often contain rich contextual clues, such as personal/family names, social context, geotags, timestamps and so on.

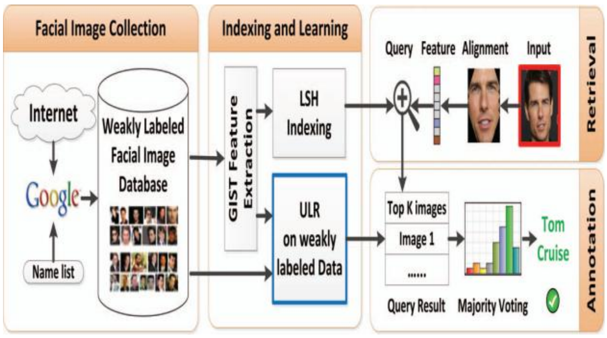
These techniques usually achieve fairly accurate annotation results, in which some techniques have been successfully deployed in commercial applications, for example, Apple iPhoto, Google Picasa, Microsoft easyAlbum and Facebook face autotagging solution. Jae young choi et al. [1] they proposed a novel collaborative framework of face recognition for improving the accuracy of face annotation. Multiple FR engines available in online social networks (OSN‟s) are used for effective FR. This paper includes two main tasks, first is the selection of expert FR engines to recognize query face images. And second is the merging of multiple FR results, generated from different FR engines, into single FR results. These works implement the viola-Jones face detection algorithm for detecting facial images in personal photos.

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, Ozkan and Duygulu [12] proposed a graph-based model for finding the densest sub-graph as the most related result. Following the graph-based approach, Le and Satoh [13] proposed a new local density score to represent the importance of each returned images, and Guillaumin et al. [14] introduced a modification to incorporate the constraint that a face is only depicted once in an image. 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 [15] to improve over the generative approach and avoid the explicit computation in graph-based approach. 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. [16] proposed a possibility model combined with a clustering algorithm to estimate the relationship between the facial images and the names in their captions.

The fifth group is about purifying web facial images, which aims to leverage noisy web facial images for face recognition applications. Usually these works are proposed as a simple preprocessing step in the whole system without adopting sophisticated techniques.

III. SEARCH BASED FACE ANNOTATION

The below fig.2 illustrates the system flow of the existing system search-based face annotation.



(a) (b) (c)

Fig.2 The system flow of the search-based face annotation scheme

Search based face annotation framework consist of the following steps.

Step 1 to step 4 are conducted before test phase of face annotation task. Step 5 and step 6 are conducted during the test phase of face annotation task.

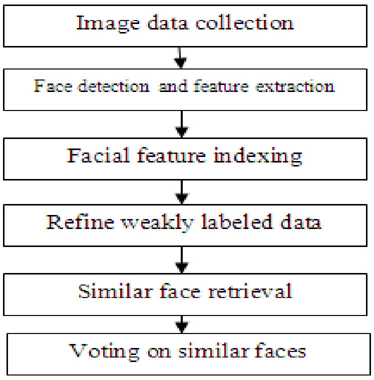


Fig. 3 Search based face annotation

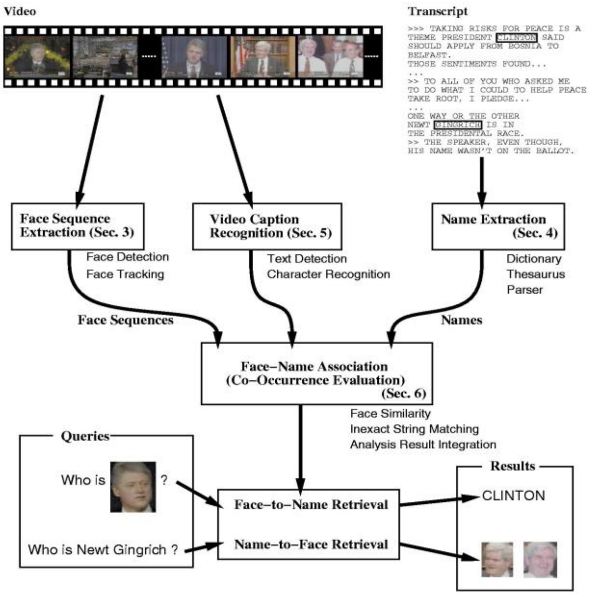
The step 1 is the data collection of facial images in which facial images are collected from the WWW by an existing web search engine i.e. Google. These facial images are often noisy, which do not always correspond to the right human name. Such kinds of web facial images with noisy names are called as weakly labeled facial image data. The step 2 is the preprocess web facial images to extract face-related information, which includes face detection and alignment, facial region extraction and facial feature representation. The step 3 is to index the extract features of the faces. The step 4 is the unsupervised learning scheme to enhance the label quality of the weakly labeled facial images. The step 5 is the process of similar face. In the step 6 the majority voting approach is applied.

This system does not effectively exploit the short list of candidate facial images problem with face recognition. Sometimes Issues of duplicate human names/labeling.

IV. PROPOSED WORK

A. Video Caption-Recognition

Attached directly to image sequences, video captions provide text information.



>>> TAKING RISKS FOR PEACE IS A
THEME PRESIDENT CLINTON SAID
SHOULD APPLY FROM BOSNIA TO
BELFAST.
THOSE SENTIMENTS FOUND...
...
>> TO ALL OF YOU WHO ASKED ME
TO DO WHAT I COULD TO HELP PEACE
TAKE ROOT, I PLEDGE...
...
ONE WAY OR THE OTHER
NEWT GINGRICH IS IN
THE PRESIDENTAL RACE.
>> THE SPEAKER, EVEN THOUGH,
HIS NAME WASN'T ON THE BALLOT.

Fig. 4. Name-It System

In many cases, they're attached to faces and usually represent a person's name. Thus, video-caption recognition provides rich information for face-name association, although not necessarily attached to all faces of persons of interest.

1. Face sequence extraction:

A video is given as input. The input video is made up of frames. Approx 20-25 frames/sec are there in a video. These frames are extracted from the video. Total frames in a

video = Frame rate * total seconds in video.

2. Face recognition:

1. Verification (one-to-one matching):

When presented with a face image of an unknown individual along with a claim of identity, ascertaining whether the individual is who he/she claims to be.

2. Identification (one-to-many matching):

Given an image of an unknown individual, determining that person's identity by comparing (possibly after encoding) that image with a database of (possibly encoded) images of known individuals.

B. Major kinds of Relationships

The three kinds of relationships are consider,

- Face-to-name resemblance

Models how likely a face should be assigned to a name based on external knowledge from image domain.

- Face-to-face constraint

Consider the factors such as background context, spatial overlap, temporal disconnectivity and visual similarity for relating faces from different frames and videos.

- **Name-to-name relationship or social relation:**

Considers the joint appearance of celebrities by leveraging social network constructed based on the co-occurrence statistics among celebrities.

The first two relationships are subjugated for labeling faces in a video, that is "within-video" face labeling. The task is to assign the names mentioned in metadata to the faces detected in a video, with the problem of missing faces and names in mind such that "null assignment" of names. Another is "between-video" naming, which performs labeling of faces on a group of videos whose celebrities fall in the same social network.

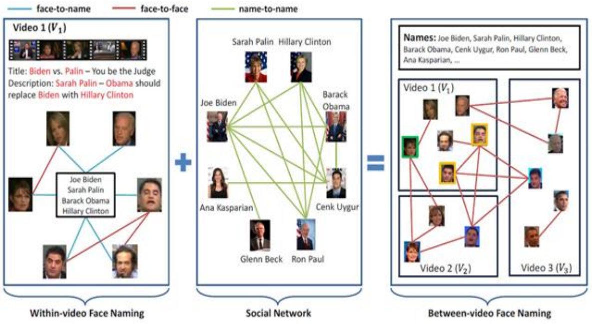


Fig. 3. Two major task of the system

Fig. 3. depicts two major tasks in this paper. Given a Web videoV1, within-video labeling constructs a graph with the names and faces in the video as vertices. Based upon the face-to name and face-to-face relationships, edges are established among the vertices for inference of face labels by CRF. The inference can be affected by situations such as there are faces whose names are not mentioned in the metadata (e.g., Cenk Uygur), and similarly names mentioned in the metadata but faces do not appear in the video (e.g., Barack Obama). Between-video face labeling, by associating V1 to a social network, crawls relevant videos (i.e., V2 andV3) and forms a larger graph composing of names and faces from multiple videos. Using social cues, additional edges modeling name-to-name relationships are also established.

As shown in the example of Fig. 3, the expanded graph has the advantages that the missing name Cenk Uygur (marked in yellow rectangle) in V1 can be propagated from V2and V3 and the corresponding faces are assigned with the name replacing the null label, while the face wrongly labeled as Hillary Clinton (marked in green rectangle) can be rectified with name-to-name relationship as well as the similar faces found inV2.

V. RELATIONSHIP MODELING

Multiple relationships are defined to characterize the sets of faces and names in the CRF.

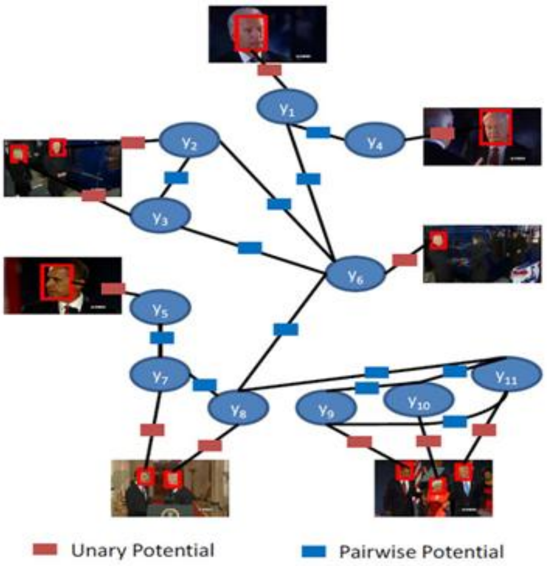


Fig. 4. Graph depicting the modeling of relationships for face naming

C. Unary Potential
It measures the like hood of faces being labeled with a name or "null".

D. Pairwise Potential
It characterizes the possible relationships between two faces.

1. Spatial Relationship
For a given two frames of different shots, the spatial locations of faces, as well as their overlapping area, give clue to the identity of face.

2. Temporal Relationship
The appearance of faces at different timestamps along the temporal axis gives clue of whether the names assigned to faces should be exclusive of each other.

3. Visual Relationship
The inference of labels based on face dissimilarity.

VI. CLUSTERING-BASED APPROXIMATION

The problem were considered as n*m, where n is the number of facial images in the retrieval database and m is the number of distinct names. For small problem, it can be efficiently solve by MGA-based algorithm. For large problem it can adapt the CDA-based algorithm. However, when n is extremely large, the CDA-based algorithm can be computationally intensive. One of the solution to it is to adapt parallel computation. But speedup of the parallel computation approach is depending on hardware capability. To advance the scalability and efficiency in algorithm here is the clustering-based approximation solution.

Clustering concept could be applied in two different levels:

A. "image-level" which is used to directly separate all the n facial images into set of clusters.
B. "name-level" which is used to first separate the m names into clusters, then to further split retrieval database into different subsets according to name-label clusters. "image-level" clusters would be more time—consuming than "name-level" because the number of facial images n is much larger than the number of names m.

VII. APPLICATIONS

Face annotation finds its application in the field of

- Achieve relatively high performance without user interaction.
- When user interaction is included, reduce to an acceptable level.
- Online photo album management and also in video domain.

VIII. CONCLUSION

This paper presents an extensive survey on face annotation techniques for web facial images found in videos. In this report we have discussed the modeling of multiple relationships using CRF for celebrity naming in the Web video domain. In view of the incomplete and noisy metadata, CRF softly encodes these relationships. It allows null assignments by considering the indecision in labeling.

REFERENCES

[1] J.Y. Choi, W.D. Neve, K.N. Plataniotis, and Y.M. Ro, "Collaborative Face Recognition for Improved Face Annotation in Personal Photo Collections Shared on Online Social Networks," IEEE Trans. Multimedia, vol. 13, no. 1, pp. 14-28, Feb. 2011.

[2] 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.

[3] X.-J. Wang, L. Zhang, F. Jing, and W.-Y. Ma, "AnnoSearch: Image Auto-Annotation by Search," Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR), pp. 1483-1490, 2006.

[4] L. Wu, S.C.H. Hoi, R. Jin, J. Zhu, and N. Yu, "Distance Metric Learning from Uncertain Side Information for Automated Photo Tagging," ACM Trans. Intelligent Systems and Technology, vol. 2, no. 2, p. 13, 2011.

[5] P. Wu, S.C.H. Hoi, P. Zhao, and Y. He, "Mining Social Images with Distance Metric Learning for Automated Image Tagging," Proc. Fourth ACM Int'l Conf. Web Search and Data Mining (WSDM '11), pp. 197-206, 2011.

[6] J. Zhu, S.C.H. Hoi, and L.V. Gool, "Unsupervised Face Alignment by Robust Nonrigid Mapping," Proc. 12th Int'l Conf. Computer Vision (ICCV), 2009.

[7] C. Siagian and L. Itti, "Rapid Biologically-Inspired Scene Classification Using Features Shared with Visual Attention," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, no. 2, pp. 300-312, Feb. 2007.

[8] W. Dong, Z. Wang, W. Josephson, M. Charikar, and K. Li, "Modeling LSH for Performance Tuning," Proc. 17th ACM Conf. Information and Knowledge Management (CIKM), pp. 669-678, 2008.

[9] W. Dong, Z. Wang, W. Josephson, M. Charikar, and K. Li, "Modeling LSH for Performance Tuning," Proc. 17th ACM Conf. Information and Knowledge Management (CIKM), pp. 669-678, 2008.

[10] P. Pham, M.-F. Moens, and T. Tuytelaars, "Naming Persons in News Video with Label Propagation," Proc. VCIDS, pp. 1528-1533, 2010.

[11] J. Tang, R. Hong, S. Yan, T.-S. Chua, G.-J. Qi, and R. Jain, "Image Annotation by KNN-Sparse Graph-Based Label Propagation over Noisily Tagged Web Images," ACM Trans. Intelligent Systems and Technology, vol. 2, pp. 14:1-14:15, 2011.

[12] D. Ozkan and P. Duygulu, "A Graph Based Approach for Naming Faces in News Photos," Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR), pp. 1477-1482, 2006.

[13] D.-D. Le and S. Satoh, "Unsupervised Face Annotation by Mining the Web," Proc. IEEE Eighth Int'l Conf. Data Mining (ICDM), pp. 383-392, 2008.

[14] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "Automatic Face Naming with Caption-Based Supervision," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2008.

[15] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "Face Recognition from Caption-Based Supervision," Int'l J. Computer Vision, vol. 96, pp. 64-82, 2011.

DOI 10.17148/IARJSET/NCIARCSE.2017.18