RESEARCH ARTICLE OPEN ACCESS

Review of Identification of Face-Name in Videos

Ankita Somani[1], Bharati Sonawane[2], Amruta Shingare[3],Nikita Gaikwad[4]
1(Department Of Computer Engineering, Vpcoe, Baramati)
2(Department Of Computer Engineering, Vpcoe, Baramati)

Abstract:

This paper explores the issue of celebrity face naming in unconstrained videos with client gave metadata. Rather than depending on exact face names for supervised learning, a rich arrangement of relationships consequently got from video content and information from picture area and social cue is utilized for unsupervised face labeling. The relationships refer to the appearances of faces under changed spatiotemporal contexts and their visual similarities. The learning incorporates Web pictures weakly labeled with celebrity names and the celebrity social networks. The relationships and learning are exquisitely encoded utilizing conditional random field(CRF) for label inference. Two versions of face annotation are considered: within-video and between-video face labeling. The previous locations the issue of incomplete and noisy labels in metadata, where invalid assignment of names is permitted an issue rarely been considered in the literature. The last further rectifies the errors in metadata, particularly to correct false names and clarify faces with missing names in the metadata of a video, by considering a gathering of socially associated videos for joint name inference.

Keywords — **Face naming, Social network, Web videos, Within-video face naming, Between-video face naming**

I.INTRODUCTION

When we see videos on web, for each person there may occur different variations in faces like hairstyle, eyebrows, wrinkles, skin tone etc. Due to these variations in faces, labeling faces in web videos is a challenging problem. To date, most search engines index these videos with user-provided text descriptions e.g., title, tag, which are noisy and incomplete. It is usual that a mentioned person may not appear in the video, and a celebrity actually appearing in a video is not mentioned. Because of low recall or low precision. Searching people-related videos may give unsatisfactory performance. Utilization of rich context information for face cannot be directly applied for domain unrestricted videos, because of the absence of context cues and prior knowledge such as family relationships for problem formulation. In this paper, leveraging on rich relationships rather than rich texts [1] in Web video domain, an algorithm based on conditional random field (CRF) [4] is proposed to address the problem of face naming.

II. EXISTING SYSTEM

The existing research efforts for face naming are mostly dedicated to the domain of Web images and constrained
videos such as TV serials, news videos and movies. These works can be broadly categorized into three groups:

***A. Model-based* :**

Model-based approaches seek to learn classifiers for face recognition. Due to the requirement of labeled samples as training examples for each face model, these approaches generally do not scale with the increase number of names. Multiple instance learning[2] investigate a more challenging problem where there is no complete knowledge on data labels available to the learning methods. Each face-name association is treated as a data instance with a binary correct or incorrect label, and a classifier is trained from

labeled data using a supervised method to predict the probability that a name is correctly associated with a face. In [3] large scale learning weakly-labeled images directly crawled from Web are used for learning of face models. It propose a system that can learn and recognize faces by combining signals from large scale weakly labeled text, image, and video collection. DeepFace [9] is the most recent work achieving great success by using deep learning methods.

It develops an effective Deep Neural Net (DNN) architecture and learning method that leverage a very large labeled dataset of faces in order to obtain a face representation that generalizes well to other datasets. It also contributes towards an effective facial alignment system based on explicit 3D modeling of faces. Deep Face is resourcefully expensive as it requires large number of training samples. Therefore, model based approaches are time consuming and expensive to collect a large amount of human-labeled training facial images.

B. Search based:

In search-based approaches there is no need for training examples since no classifier will be explicitly trained. Search-based approaches mine the names from the retrieved examples deemed to be similar to the query faces. An effective Unsupervised Label Refinement (ULR) [5] approach uses refining of the labels of web facial images by exploring machine learning techniques. It develop effective optimization algorithms to solve the large scale learning tasks efficiently.In [6], the local coordinate coding (LCC) is applied for an effective Weak Label Regularized Local Coordinate Coding (WLRLCC) technique, which exploits the local coordinate coding principle in learning sparse features, and meanwhile employs the graph-based weak label regularization principle to enhance the weak labels of the short list of similar facial images. The main challenge for this line of approaches is to conquer the problem of noisy labels. A multimodal learning scheme[8] investigates a search-based face annotation (SBFA) paradigm for mining web facial images. The idea of SBFA is to first search for top-n similar facial images from a web facial image database and then exploit these top-ranked similar facial images and their weak labels for naming the query facial image. Search based approaches assume each name corresponds to a unique single person, which makes the problem more clearly. However, this is not always true for real-life scenarios. For example, it is possible that two persons have the same name or one person may have multiple names.

C. Clustering-based

These approaches generally perform well when there are only a few name candidates to be considered for a face.

Graph-based clustering (GC) [7] consider two scenarios of naming people in databases of news photos with captions: (i) Finding faces of a single person, and (ii) Assigning names to all faces. System combine an initial text-based step, that restricts the name assigned to a face to the set of names appearing in the caption, with a second step that analyses visual features of faces. Face-name association by commute distance (FACD) [8] has one off-line index stage and one on-line match stage. In the off-line stage, a Name-face index structure is established for efficient face retrieval. In the on-line stage, each request image-caption item is processed independently. Specifically, it builds a unified graph for each item and measure the face-name relationship via commute distance on the graph. Constraint Gaussian mixture models (CGMM) [11] learns a Gaussian mixture model for each name. The learning iterates between assigning faces to the best possible model (E-step) and updating of model parameters (M-step). A null category for dealing with missing names problem is also learnt by treating all the faces as a mixture model. Conditional Random Field(CRF)[4] presents iterative parameter estimation algorithms for CRF and compare the performance of the resulting models to Hidden Markov Models (HMMs) and Maximum Entropy Markov models (MEMMs) on synthetic and natural-language data.

III. PROPOSED APPROCH

The proposed technique, leveraging on rich relationships rather than rich texts [1] in web video domain, an algorithm based on conditional random field (CRF) [4] is proposed to address the problem of face naming.

Systems consider the following three major types of relationships:

- *Face-to-Name:* A degree to which a face should be assigned to a name on the basis of external knowledge from image domain.
- *Face-to-Face:* It considers factors such as spatial overlap, temporal dis-connectivity, background context and visual similarity for relating faces from different frames and videos.
- *Name-to-Name:* It considers the joint appearance of persons by taking advantage of social network constructed based on the co-occurrence statistics among persons.

Two versions of face annotation are considered: within-video and between -video face labeling.

A. Within-video Naming:

Fig. 1 depicts two major tasks in this paper. Given a Web video V1, "within-video" labelling 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 between the vertices for inference of face labels by CRF, with the problem of missing faces and names in mind such that null assignment of names is allowed. 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).

B. Between-video Face Naming:

The social relationship extends graph of within-video to between-video naming, by performing labeling of faces on a group of videos, where persons fall in the same social network. "Between-video" face labeling, by associating V1 to a social network, collect relevant videos (i.e., V2 and V3) 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. 1, the expanded graph has the advantages that the missing name "Cenk Uygur" (marked in yellow rectangle) in V1 can be propagated from V2 and 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 in V2.

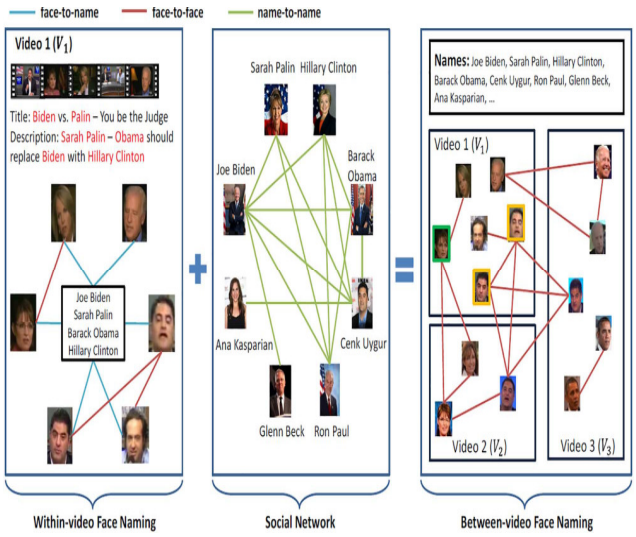

Fig.1. Within-video naming constructs a graph modeling the face-to-name and face-to-face relationships among the faces and names found in a video (v1). By social network, between-video labeling expands the graph by connecting to the graphs of two other videos (v2 and v3) that share social relations. The expanded graph is additionally modelled with name-to-name relationship inferred from the social network [1].

IV. CONCLUSION

A review of techniques of Identification of Face-Name in Videos.The proposed method uses rich relationships rather than rich texts. It improves the recent graph-based approaches. It further rectifies the errors in metadata, particularly to correct false names and clarify faces with missing names in the metadata of a video, by considering a gathering of socially associated videos for joint name inference. The consideration of between-video relationships also results in performance boost, mostly attributed to the capability of rectifying the errors due to missing names and persons.

REFERENCES

1. Lei Pang, Chong-Wah Ngo , "Unsupervised Celebrity Face Naming in Web Videos," IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 17, NO. 6, JUNE 2015.

2. J. Yang, R. Yan, and A. G. Hauptmann,"Multiple instance learning for labeling faces in broadcasting news video," in Proc. ACM Int. Conf. Multimedia, 2005, pp. 3140.

3. M. Zhao, J. Yagnik, H. Adam, and D. Bau, "Large scale learning and recognition of faces in web videos, " in Proc. Int. Conf. Automat. Face Gesture Recog., 2008, pp. 17.

4. J. D. La_erty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: probabilistic models for segmenting and labeling sequence data, " in Proc. Int. Conf.
Mach. Learn., 2001, pp. 282289.

5. D. Y. Wang, S. Hoi, Y. He, and J. K. Zhu, "Mining weakly labeled web facial images for search-based face annotation, " IEEE Trans. Knowl.Data Eng., vol. 26, no. 1,

6. D. Y.Wang, S. C. Hoi, Y. He, J. K. Zhu, T. Mei, and J. B. Luo, "Retrieval-based face annotation by weak label regularized local coordinate coding, " IEEE Trans. Pattern

7. M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "Automatic face naming with caption-based supervision," in Proc. IEEE Comput. Vis. Pattern Recog., Jun.

8. JBu et al., "Unsupervised face-name association via commute distance," in Proc. ACM Int. Conf.Multimedia,2012,pp.219-228.

9. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, ”DeepFace: Closing the gap to human-level performance in face verification, ” in Proc. IEEE Comput. Vis. Pattern Recog., Jun. 2014, pp. 17011708

10. D. Y. Wang, S. C. Hoi, P. C. Wu, J. K. Zhu, Y. He, and C.Y.
Miao,”Learning to Name Faces: A Multimodal Learning
Scheme for Search-Based Face Annotation,” in Proc. ACM
Conf. Res. Develop. Inf. Retrieval, 2013, pp. 443452.

11. T. Berg, A. Berg, J. Edwards, and D. Forsyth, ”Whos in
the picture?,” in Proc. Neural Inf. Process. Syst., 2005, pp.
137144.