A SURVEY ON NAMING FACES USING ANNOTATIONS BASED ON EXTERNAL KNOWLEDGE FROM VIDEOS

Tushar Atakari, Rishabh Kadam, Shrikrushna Kamble, Akshay Mahajan

1. *Tushar Atakari, BE(IT),DYPIET(pimpri),Maharashtra,India*
2. *Rishabh Kadam, BE(IT),DYPIET(pimpri),Maharashtra,India*
3. *Shrikrushna Kamble, BE(IT),DYPIET(pimpri),Maharashtra,India*
4. *Akshay Mahajan, BE(IT),DYPIET(pimpri),Maharashtra,India*

ABSTRACT

This paper examine the difficulty of celebrity face naming in unrestricted videos with user-provided metadata. Instead of relying on precise face labels for supervised knowledge, a rich set of associations robotically derived from video content and information from image domain and social cues is leveraged for unverified face labelling. The relations refer to the appearances of faces under dissimilar spatio-temporal contexts and their visual similarities. The knowledge includes Web images weakly tagged with celebrity names and the celebrity social networks. The associations and information are elegantly encoded using conditional random field (CRF) for label inference. Two versions of face annotation are considered: within-video and between-video face labelling. The previous addresses the problem of incomplete and noisy labels in metadata, where null assignment of names is allowed—a difficulty seldom been measured in the literature. The concluding further rectifies the errors in metadata, specifically to correct false labels and annotate faces with missing names in the metadata of a video, by allowing for a group of socially connected videos for joint label inference.

Keywords: Celebrity face naming, social network, unconstrained web videos, unsupervised.

---

1. INTRODUCTION

Due to the reputation of various digital cameras and the quick expansion of social media tools for internet-based photo-video sharing, recent years have witnessed an raised of the number of digital photos captured and stored by consumers. A huge part of photos/videos shared by users on the Internet are human facial images. Some of these facial images are tagged with names, but many of them are not tagged properly.

This has encouraged the learn of auto face annotation, an chief method that aims to annotate facial images automatically. Auto face annotation can be useful to many real world applications. For example, with auto face annotation techniques, online photo-sharing sites (e.g., Facebook) can automatically annotate user's uploaded photos to make easy online photo search and management. Besides, face annotation can also be functional in news video area to distinguish important persons appeared in the videos to make possible news video retrieval and summarization tasks.

Classical face annotation approaches are often treated as an comprehensive face recognition problem, where dissimilar categorization models are trained from a compilation of well-labelled facial images by employing the supervised or semi-supervised machine learning techniques. However, the "model-based face annotation" techniques are limited in several aspects. First, it is usually slow and high-priced to collect a large amount of human labelled training facial images. Second, it is naturally difficult to simplify the models when new training data or new persons are added, in which an exhaustive retraining process is usually required.

Recently, some rising studies have attempted to discover a talented search-based annotation paradigm for facial image annotation by mining Web (WWW), where a enormous numeral of weakly labelled facial

images are freely available. Instead of training explicit categorization 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 developing content-based image retrieval(CBIR) techniques [8], [9] in mining large 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 [10], [11], [12] for generic image annotations.

2. REVIEW OF EXISTING WORK

This section reviews the main existing work found in the scientific literature that applies Video Live Streaming over Peer to Peer Network.

The unusual raise of video on the web and the growing sparseness of meta-information connected with it forces us to look for signals from the video content for search/information recovery and browsing based corpus exploration. A large chunk of users' searching/browsing patterns are centered around folks present in the video. It is difficult due to a) the lack of labelled data for such a huge set of people and b) the great variation of pose / illumination / expression / age / occlusion /quality etc. in the target corpus.

We recommend a system that can learn and distinguish faces by combining signals from large scale weakly labelled text, image, and video corpora. First, consistency education is recommended to create face models for popular persons. We use the text-image co-occurrence on the web as a weak signal of significance and learn the set of consistent face models from this very large and noisy training set. Second, well-organized and accurate face detection and face tracking is applied. Last, the key faces in every face track is choose by clustering to get packed together and strong representation. The face tracks are additional clustered to get more representative key faces and eliminate duplicate key faces. For each cluster of face tracks, a mixture of widely held voting and probabilistic voting is done with the automatically learned models. The efficacy of our framework is demonstrated by results on image and video corpora, in which we can achieve 92.68% in 37 million[1].

The task of unverified face-name association has received a significant interests in multimedia and information recovery communities. It is fairly different with the generic facial image annotation problem because of its unverified and ambiguous task properties. Specifically, the task of face-name association should obey the following three constraints:
(1) a face can only be allocated to a name come into view in its connected caption or to *null* ;
(2) A name can be allocated to at most one face.
(3) A face can be allocated to at most one name.
Many conventional methods have been recommended to attempt this task while suffering from some ordinary problems, In this paper, we aim a novel framework named face-name association via commute distance (FACD), which judges face-name and face-null assignments under a unified framework via commute distance (CD) algorithm. Then, to additional speed up the on-line processing, we recommend a novel anchor-based commute distance (ACD) algorithm whose idea is using the anchor point representation structure to accelerate the Eigen decay of the adjacency matrix of a graph. Systematic experimentation results on a big scale and real image-caption database with a total of 194,046 detected faces and 244,725 names show that our recommended approach out performs many state-of-the art methods in performance. Our agenda is appropriate for a large scale and real-time system[2].

Huge video collection consisting of news programs, dramas, movies, and web videos (e.g., YouTube) are available in our daily life. In all these videos, human is typically one of the most significant subjects. Using state-of-the-art methods, we can capably sense and track faces in the videos. In order to arrange large-scale face tracks, containing series of (detected) successive faces in the videos, we recommend an efficient way to recover human face tracks by means of bag-of-faces sparse representation. Using the recommended process, a face track is encoded as a single bag-of-faces sparse representation and therefore allowing capable indexing method to handle large-scale data. To additional think the probable variations in face tracks, we simplify our method to discover lots of sparse representations, in an unverified manner, to stand for a bag of faces and equilibrium the trade-off among performance and retrieval time. Experimental results on two real-world (million-scale) datasets confirm that the recommended methods achieve significant performance gains compared to different state-of-the-art methods[3].

Associating faces present in Web videos with names presented in the nearby background is an significant task in lots of application. However, the difficulty is not well investigated mainly under large-scale sensible scenario, mainly due to the shortage of dataset constructed in such condition. In this paper, we set up a Web video dataset of celebrities, named Web V-Cele, for name-face relationship. The dataset consists of 75 073 Internet videos of over 4 000 hours, covering 2 427 celebrities and 649 001 faces. This is, to our information, the most inclusive dataset for this problem. We explain the details of dataset building, discuss some interesting findings by analysing this dataset like celebrity community discovery, and provide new results of name-face association by means of five existing techniques. We also draw round important and demanding research problems that could be investigated in the future[4].

We think two scenarios of naming people in databases of news photos with captions:
(I) finding faces of a single person,
(ii) Assigning names to all faces.

We unite an initial text-based step, that limit the name allocated to a face to the set of names appearing in the caption, with a second step that analyses visual features of faces. By searching for groups of highly similar faces that can be linked with a name, the results of purely text-based look for can be deeply ameliorated. We get better a recent graph-based approach, in which nodes correspond to faces and edges connect highly similar faces.

We bring in constraints when optimizing the purpose function, and recommend improvement in the low-level methods used to build the graphs. Furthermore, we simplify the graph based approach to face naming in the full data set. In this multi-person naming case the optimization quickly becomes computationally demanding, and we present an significant speed-up using graph-flows to calculate the optimal name assignments in documents. Generative models have previously been recommended to solve the multi-person naming task. We evaluate the generative and graph-based methods in both scenarios, and find significantly better performance using the graph-based methods in both cases [5].

Automated face annotation aims to automatically detect human faces from a photo and extra name the faces with the equivalent human names. In this paper, we tackle this open problem by investigating a search-based face annotation (SBFA) paradigm for mining large amounts of web facial images freely available on the WWW. Given a query facial image for annotation, the thought of SBFA is to first search for top-n similar facial images from a web facial image database and then use these top-ranked similar facial images and their weak labels for naming the query facial image. To fully mine those information, this paper recommends a novel framework of Learning to Name Faces (L2NF) – a unified multimodal learning approach for search-based face annotation, which consists of the following major components:

(i) We enhance the weak labels of top-ranked similar images by exploiting the “label smoothness" assumption;
(ii) We build the multimodal representations of a facial image by extracting different types of features;
(iii) We optimize the distance measure for each type of features using distance metric learning techniques;
(iv) We study the optimal mixture of multiple modalities for annotation through a learning to rank scheme. We carry out a set of extensive empirical studies on two real-world facial image databases, in which encouraging results show that the recommended algorithms significantly boost the naming correctness of search-based face annotation task[6].

In modern face recognition, the conventional pipeline consists of four stages:
detect ⇒ align ⇒ represent ⇒ classify. We revisit both the alignment step and the illustration step by employing explicit 3D face modelling in order to be significant a piecewise affine transformation, and derive a face illustration from a nine-layer *deep* neural network. This deep system involves more than 120 million parameters using several locally connected layers without weight sharing, rather than the standard convolutional layers. Thus we trained it on the largest facial dataset to-date, an identity labelled dataset of four million facial images belonging to more than 4,000 identities. The learned representations coupling the correct model-based alignment with the great facial database generalize remarkably well to faces in unconstrained environments, even with a simple classifier. Our process reaches an accuracy of 97.35% on the Labelled Faces in the Wild (LFW) dataset, reducing the error of the current state of the art by more than 27%, closely approaching human-level performance [7].

We explain a probabilistic method for identifying characters in TV series or movies. We aim at labelling every character look, and not only those where a face can be detected. Consequently, our basic unit of appearance is a person track (as opposed to a face track). We model each TV series episode as a Markov Random Field, integrating face recognition, clothing appearance, speaker recognition and contextual constraints in a probabilistic manner. The recognition task is then formulated as an energy minimization problem. In order to identify tracks lacking faces, we study clothing models by adapting available face recognition results. Within a scene, as pointed to by prior analysis of the temporal structure of the TV series, clothing features are combined by agglomerative clustering. We price our approach on the first 6 episodes of The Big Bang Theory and achieve an absolute improvement of 20% for person identification and 12% for face recognition[8].

In this paper, we examine a search-based face annotation framework by mining weakly labelled facial images that are freely available on the internet. A key component of such a search-based annotation paradigm is to build a database of facial images with accurate labels. This is however challenging since facial images on the WWW are often noisy and incomplete. To get better the label quality of raw web facial images, we recommend an effective Unverified Label Refinement (ULR) approach for refining the labels of web facial images by exploring machine learning techniques. We expand effective optimization algorithms to solve the large-scale knowledge tasks efficiently, and conduct an extensive empirical study on a web facial image database with 400 persons and 40,000 web facial images. Encouraging results showed that the recommended ULR method can significantly boost the performance of the promising search based face annotation scheme [9].

Retrieval-based face annotation is a promising paradigm in mining massive web facial images for automated face annotation. Such an annotation paradigm usually encounters two key challenges. The first challenge is how to efficiently retrieve a short list of most similar facial images from facial image databases, and the second challenge is how to efficiently perform annotation by exploiting these alike facial images and their weak labels which are often noisy and incomplete. In this paper, we mainly focus on tackling the second challenge of the retrieval-based face annotation paradigm. In particular, we recommend an effective Weak Label Regularized Local Coordinate Coding (WLRLCC) technique, which used 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. We present an efficient optimization algorithm to solve the WLRLCC task, and expand an effective sparse reconstruction scheme to perform the final face name annotation. We conduct a set of extensive empirical studies on a large-scale facial image database with a total of 6, 000 persons and over 600, 000 web facial images, in which encouraging results show that the recommended WLRLCC algorithm significantly boosts the performance of the regular retrieval based face annotation approaches[10].

Labelling faces in news video with their names is an fascinating research problem which was previously solved using supervised methods that demand significant user e® or its on labelling training data. In this paper, we examine a more challenging setting of the problem where there is no total in order on data labels. Specially, by using the uniqueness of a face's name, we make the problem as a special multi-instance learning (MIL) problem, namely exclusive MIL or eMIL problem, so that it can be attempt by a model skilled with partial labelling information as the anonymity judgment of faces, which requires less user effort to collect. We suggest two discriminative probabilistic learning methods named Exclusive Density (ED) and Iterative ED for eMIL problems. Experiments on the face labelling problem shows that the performance of the recommended approaches are superior to the traditional MIL algorithms and close to the performance achieved by supervised methods trained with complete data labels[11] .

Face annotation in images and videos enjoys a lot of potential applications in multimedia in order retrieval. Face annotation usually requires many training data labelled by hand in order to build effective classifiers. This is mainly demanding when annotating faces on large-scale collections of media data, in which enormous labelling efforts would be very expensive. As a consequence, traditional supervised face annotation methods often experience from insufficient training data. To attack this confront, in this paper, we recommend a novel Transductive Kernel Fisher Discriminant (TKFD) scheme for face annotation, which outperforms established supervised annotation methods with few training data. The major idea of our approach is to solve the Fisher's discriminant using deformed kernels incorporating the information of both labelled and unlabeled data. To appraise the effectiveness of our method, we have conducted extensive experiments on three types of multimedia test beds: the FRGC benchmark face dataset, the Yahoo! web image compilation, and the TRECVID

video data collection. The experimental results show that our TKFD algorithm is more successful than traditional supervised approaches, especially when there are very few training data [12].

CONCLUSION AND FUTURE WORK

We have presented the modelling 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 while allowing null assignments by considering the uncertainty in labelling. Experimental results basically show that these nice properties lead to performance superiority over several existing approaches. The consideration of between video relationships also results in further performance boost, mostly attributed to the capability of rectifying the errors due to missing names and persons. The price of improvement, never the less, also comes along with raise in processing time and the number of false positives. Fortunately, the proposals of leveraging social relation and joint labelling by sequential video processing still make CRF scalable in terms of speed and memory efficiency. While the overall performance of the recommended approach is encouraging, the effectiveness is still limited by facial feature similarity, which is used in the unary energy term and pair wise visual relationship. With the recent advancement in facial feature representations such as Deep Face [23] and face track [33], we plan to investigate the effectiveness of incorporating these representations into the recommended CRF framework in the near future.

REFERENCES

[1] Zhangyu Chang and S.-H. Gary Chan, Senior Member, IEEE, "Bucket-Filling: An Asymptotically Optimal Video-on- Demand Network With Source Coding," IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 17, NO. 5, MAY 2015.

[2] Haiying Shen, Senior Member, IEEE, Member, ACM, Yuhua Lin, and Jin Li, Fellow, IEEE, "A Social-Network-Aided Efficient Peer-to-Peer Live Streaming System",IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 23, NO. 3, JUNE 2015.

[3] Min Yang ; Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook ; Yuanyuan Yang, "Applying Network Coding to Peer-to-Peer File Sharing", IEEE Transactions on Computers Volume:63 Issue:8 Aug. 2014.

[4] H. Shen, Z. Li, and J. Li, "A DHT-aided chunk-driven overlay for scalable and efficient peer-to-peer live streaming," IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 11, pp. 2125- 2137, Nov. 2012.

[5] Nicolas Kourtellis ; Department of Computer Science and Engineering University of South Florida, Tampa, FL, USA ; Adriana Iamnitchi, " Inferring peer centrality in socially- informed peer-to-peer systems", Peer-to-Peer Computing (P2P), 2011 IEEE International Conference
.

[6] Y. Liu, "Delay bounds of chunk-based peer-to-peer video streaming," IEEE/ACM Trans. Netw., vol. 18, no. 4, pp. 1195-1206, Aug. 2010.

[7] D. Wu, Y. Liu, and K. Ross, "Modeling and analysis of multichannel P2P live video systems," IEEE/ACM Trans. Netw., vol. 18, no. 4, pp. 1248-1260, Aug. 2010.

[8] F. Ramos, J. Crowcroft, R. Gibbens, P. Rodriguez, and I. White, "Channelsmuring: Minimising channel switching Delay in IPTV distribution networks," in Proc. ICME, 2010, pp.1327-1332.

[9] J. Mol, A. Bakker, J. Pouwelse, D. Epema, and H. Sips, "The design and deployment of a bittorrent live video streaming solution," in Proc. ISM, 2009,pp. 342-349.

[10] M. L. Xu, and B. Ramamurthy, "A flexible divide-and-conquer protocol for multi-view peer-to-peer live streaming," in Proc. P2P, 2009, pp. 291-300.

[11] A. Kermarrec, E. Merrer, Y. Liu, and G. Simon, "Surfing peer-to-peer IPTV: Distributed channel switching," in Proc. Euro-Par, 2009, pp. 574-586.

[12] X. Cheng, C. Dale, and J. Liu, "Statistics and social network of YouTube videos," in Proc.IWQoS, 2008, pp. 229-238.

[13] Y. Guo, C. Liang, and Y. Liu, "AQCS: Adaptive queue-based chunk scheduling for P2P live streaming," in Proc. IFIP Netw., 2008, pp. 433-444.

[14] F. Picconi and L. Massoulie, "Is there a future for mesh-based live video streaming?," in Proc. P2P, 2008, pp. 289-298.

[15] C. Wu, B. Li, and S. Zhao, "Multi channel live P2P streaming:Refocusing on servers," in Proc. IEEE INFOCOM, 2008, pp. 2029-2037.

[16] M. Wang, L. Xu, and B. Ramamurthy, "Channel-aware peer selection in multi-view peer- to-peer multimedia streaming," in Proc. ICCCN, 2008, pp. 1-6.

Tushar Ashok Atakari, received the Diploma in Information Technology from Samarth Polytechnic ,Belhe(Bangarwadi) in 2014 and currently pursuing his B.E degree in Information Technology from Dr. D. Y. Patil Institutes of Engineering and Technology, Pimpri, Pune(Maharashtra). He is good in Programming.

Kadam Rishabh Jayant,currently pursuing his B.E degree in Information Technology from Dr. D. Y. Patil Institutes of Engineering and Technology, Pimpri, Pune(Maharashtra).

Shrikrushna Ashok Kamble, received the Diploma in Information Technology from DR. R. N. Lahoti Polytechnic, Sultanpur in 2014 and currently pursuing his B.E degree in Information Technology from Dr. D. Y. Patil Institutes of Engineering and Technology, Pimpri, Pune(Maharashtra). He is good in Programming.

Mahajan Akshay Vijay, currently pursuing his B.E degree in Information Technology from Dr. D. Y. Patil Institutes of Engineering and Technology, Pimpri, Pune(Maharashtra)..