International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

Mining Weakly Labeled Web Facial Images for Search-Based Face Annotation

Dipak D. Tayade[1], Nilesh Y Chaudhari[2]

[1, 2]North Maharashtra University, Godavari Foundations College of Engineering and Technology, Jalgaon, Maharashtra, India

Abstract: ***In this paper, we present an automatic web image and video mining framework with the ultimate goal of building a universal human age estimator based on facial information, which is applicable to all ethnic groups and various image qualities. On one hand, a large (391 k) yet noisy human aging image database is collected from Flickr and Google Image using a set of human age-related text queries. Multiple human face detectors based on distinctive techniques are adopted for noise-prune face detection. For each image, the detected faces with high detection confidences constitute a bag of face instances.***

Keywords: Image processing, face annotation, weakly labeled facial images, internet vision, Age estimation

1. Introduction

IMAGE-BASED human age estimation has wide potential applications, e.g., demographic data collection for supermarkets or other public areas, age-specific human computer interfaces, age-oriented commercial advertisement, and human identification based on old ID-photos. The previous research for human age estimation can be roughly divided into two categories according to whether the age estimation task is considered as a regression problem or a multi-class classification problem.

The explosive increasing of online sharing media such as image and video sharing websites, e.g., Flickr, Picasa, and image search engines, e.g., Google Image, has shed light on obtaining a large number of training data (e.g., images and videos) for general visual learning tasks.

Note that known as web mining, utilizing the web resources such as images, videos, personal blogs as well as their corresponding tags, surrounding texts and meta-data, have been recently utilized for various computer vision and multimedia tasks. Yanai and Barnard proposed a web image mining method for discriminative visual concept selection. Zheng et al. leveraged the vast amount of multimedia data on the web to build a world-scale landmark recognition engine. Ji et al. [19] reported a city landmarks discovery and personalized tourist suggestion system by mining the images automatically crawled from online sharing personal blogs.

Auto face annotation can be 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 photo 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 idea retrieval and summarization tasks [2], [3]. Classical face annotation approaches are often treated as an extended face recognition problem, where different classification models are trained from a collection of well labeled facial images by employing the supervised or semi-supervised machine learning techniques [2], [4], [5], [6], [7]. However, the "model-based face annotation" techniques are limited in several aspects.

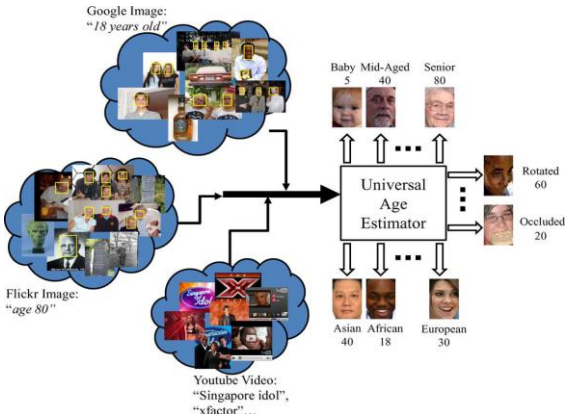

Figure 1: Illustration of the purpose of this study, i.e., to utilize web image and online video resources for learning a universal age estimator.

Noisy image and label filtering: First, we propose to conduct parallel face detection based on multiple face detectors for improving the probability to obtain well-aligned face instances for each image, and then the overlapping face instances from distinct detectors are retained as good samples for model training. Then, principal component analysis is applied for each age label and those face instances with large reconstruction errors are filtered out.

Robust multi-instance regression: Given a set of training face images with multiple face instances within each image as well as a set of label-consistent face pairs, we formulate our task as a specific multi-instance learning problem (e.g., regression) with extra constraints. As the previous multi-instance learning algorithms cannot be directly applied for this problem since there may exist noisy image labels for the training data, in this work, we present a robust multi-instance learning algorithm to tackle this problem with the awareness of label outliers. Our formulation also absorbs the label-consistence regularization for the tracked face pairs. Note that multi-instance learning is a widely studied research topic in the past few years. Keeler et al. first proposed the

multi-instance learning concept when dealing with the hand-printed numerals detection problem, motivated by the observation that there might exist more than one numeral in a single image. After that, many researchers proposed various related schemes such as DD, EM-DD, and citation-k NN to tackle this problem. Also the multi-instance learning concept was incorporated into both boosting and support vector machine algorithms, yielding the so-called MIL-boosting and MIL-SVM algorithms. There also exist several algorithms proposed for the multiple instances multiple label learning problems. Besides these classification problems, recently Ray and Page [27] proposed a multi-instance regression framework to deal with the regression problem, which is the most related work with ours. This algorithm does not consider the noisy label issue, and therefore, the algorithmic robustness cannot be guaranteed.

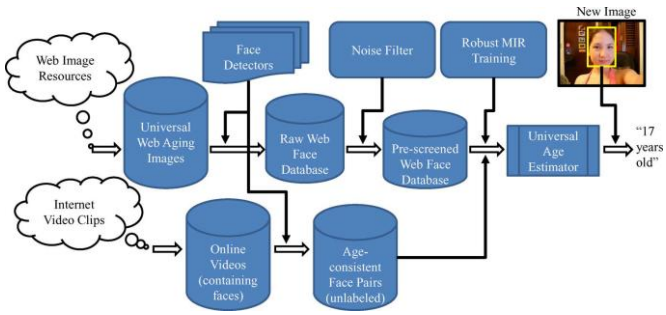

Figure 2: System overview for learning universal age estimator based on automatic web image and online video mining

The main contributions of this work are threefold: 1) we present a web image and video mining scheme to harness the Internet images and videos for collecting a large and diverse face database with nearly-correct age labels and age-consistent face pairs, and then use it for learning universal human age estimator; 2) we propose a novel learning algorithm to robustly derive a human age estimator based on images with multiple face instances and possibly noisy labels, under the constraints of age-consistent face pairs; and 3) we develop a fully automatic system for automatic training image/video collection, age repressor learning, and age estimation, which does not rely on any kind of human interactions and is thus of great potential in real scenarios. A system overview is illustrated in Fig. 2. The rest of this paper is organized as follows. Section II introduces the Internet aging image collecting and noise filtering process. The robust multi-instance regression algorithm is elaborated in Section III and the experimental results are demonstrated in Section IV. Section V concludes this paper along with discussion of future work.

2. Face Recognition Algorithm

Principal Component Analysis (PCA): PCA also known as Karhunen-Loeve method is one of the popular methods for feature selection and dimension reduction. Recognition of human faces using PCA was first done by Turk and Pent land and reconstruction of human faces was done by Kirby and Sirovich. The recognition method, known as Eigen face method defines a feature space which reduces the dimensionality of the original data space. This reduced data space is used for recognition. But poor discriminating power within the class and large computation are the well-known common problems in PCA method. This limitation is overcome by Linear Discriminant Analysis (LDA). LDA is the most dominant algorithms for feature selection in appearance based methods. But many LDA based face recognition system first used PCA to reduce dimensions and then LDA issued to maximize the discriminating power of feature selection. The reason is that LDA has the small sample size problem in which dataset selected should have larger samples per class for good discriminating features extraction. Thus implementing LDA directly resulted in poor extraction of discriminating features. In the proposed method [10] Gabor filter is used to filter frontal face images and PCA is used to reduce the dimension of filtered feature vectors and then LDA is used for feature extraction. The performances of appearance based statistical methods such as PCA, DA and ICA are tested and compared for the recognition of colored faces images in [11]. PCA is better than LDA and ICA under different illumination variations but LDA is better than ICA. LDA is more sensitive than PCA and ICA on partial occlusions, but PCA is less sensitive to atrial occlusions compared to LDA and ICA. PCA s used as a dimension reduction technique in [12] and for modeling expression deformations in [13]. A recursive algorithm for calculating the discriminant features of PCA-LDA procedure is introduced in [14]. This method concentrates on challenging issue of computing discriminating vectors from an incrementally arriving high dimensional data stream without computing the corresponding covariance matrix and without knowing the data in advance.

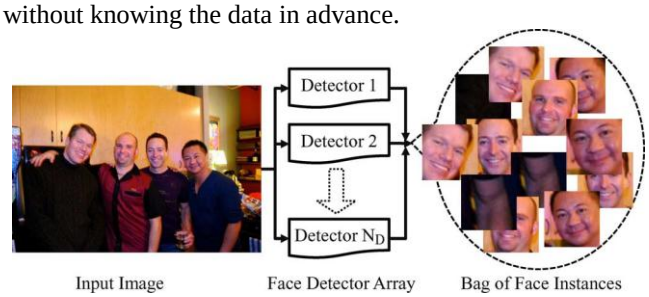

Figure 3: Exemplary result from parallel face detection

Recent years have witnessed an explosion of social media content shared online, e.g., Flickr. For these community-contributed media repositories, the users may upload personal multimedia data with informative titles and annotate them with descriptive tags. For those human face-related images, the human age information is often naturally involved within the titles, performances, but generally no detector can guarantee to be perfect. However, the parallel detection results from several face detectors can provide multiple and complimentary detection results for each image, which actually simulate the multi-instance learning scenario. We thus adopt this scheme to use multiple variations (e.g., two) of Ad boost-based face detectors for detecting all possible faces for each image. The diagram of a parallel face detection scheme is illustrated in Fig. 3.

Figure 4 shows several typical samples before the pre-screening step displays the sample face instances automatically cropped from the face image database before and after pre-screening. Several conclusions can be drawn

from these observations:

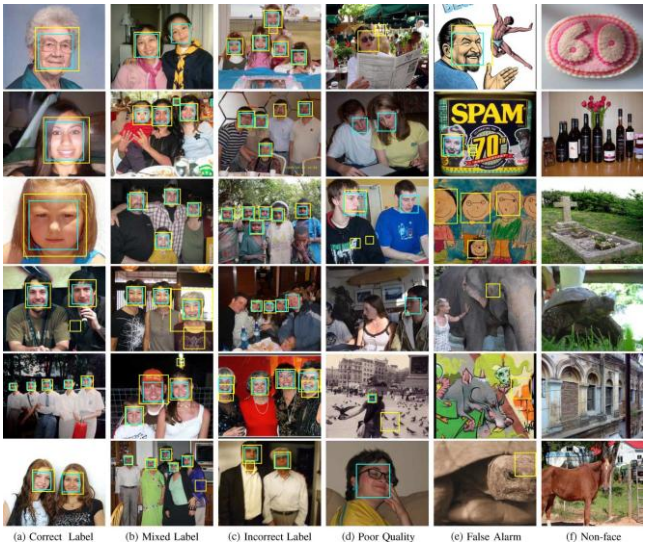

Figure 4: Some sample images from the raw face database with detected face regions. Each column denotes a type of detection results, including (from left to right): (a) all the face instances (single or multiple) within the image inherit the bag age label; (b) part of the face instances inherit the bag age label and other detected face instances correspond to other ages (noisy instances); (c) the bag age label is incorrect (the age labels for the images are 20, 10, 50, 60, 20, 20 from top to bottom); (d) poor quality face instances due to rotation, illumination variation, occlusion, or photo fadedness; (e) images contain false detections; (f) age-relevant images which however contain no face instances. Note that different colors of the detection rectangles indicate the results from different detectors.

1) In most raw images, multiple face instances are cropped due to the multiple detector process, which essentially leads to a multi-instance problem for learning the universal age estimator.
2) We could observe that for some images, the bag age label is incorrect. In this case, the original multiple instance learning algorithms is easy to fail. Therefore, a robust multi instance earning algorithm, which can handle noisy labels, is necessary for pursuing a universal age estimator.
3) There are also many poor quality faces and false alarm detections output from the face detectors. Poor quality faces include those non-frontal faces and occluded faces. Most of these inappropriate detections may be pre-screened out by the multiple detector strategy. This results in every clean database containing face instances only.
4) A small portion of the filtered instances are true faces. These true faces are also removed as they are detected by only one face detector.
5) There exist a significant number of non-face images related to the age keywords, e.g., pet, old building, wine, birthday cake, and tomb, which are easily filtered out by the face detectors.

An illustration of the age statistics for the database before and after pre-screening is shown in Figure 5.

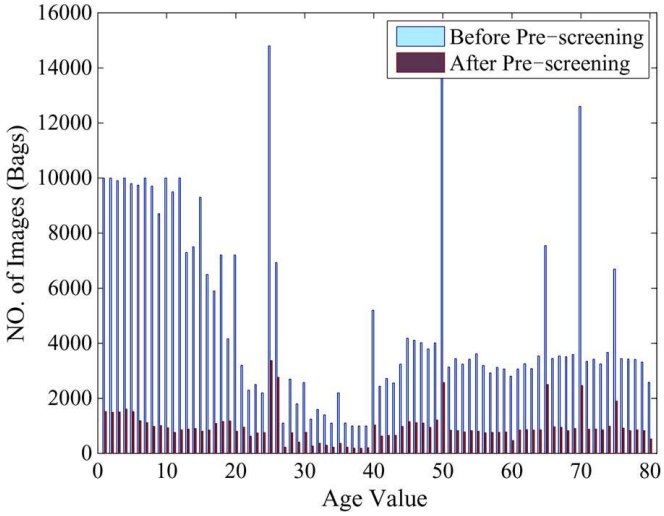

Figure 5: Age label statistics of the downloaded images before (left light color bars) and after (right dark color bars) pre-screening.

3. Label Refinement on artificial database

In this experiment, we aim to evaluate the label refinement performance of different algorithms. We built an artificial dataset that consists of 9 classes (persons) in 2-dimensional space with 20 samples for each class. To introduce noise into the label matrix, we randomly mislabeled half of the whole dataset. All the data points are illustrated in Fig. 4(a), and the original noisy label matrix is shown as the leftmost one in Fig. 4(b). Given the dataset and the noisy label matrix, we computed the enhanced label matrixes using the four algorithms mentioned in Section 2 (see Fig. 4(b)). Several observations can be drawn from the above results: first, the MKL and CL algorithms work well for the classes with less noise (e.g. Person 1 and Person 9), but they fail for number with respect to their sub problems to 30.

4. Equations

Finally, the below algorithm summarizes the optimization progress.

Algorithm 1: Multi-step Gradient Algorithm for ULR
Input: $Q \in R(n\cdot m)\times(n\cdot m)$, $c \in Rn\cdot m$, $t \in R$
Output: $x\star$
1 begin
2 $\alpha 0 = 1$;$k = 1$; $z(0) = x(0) = x(-1) = 0$;
3 repeat
4 Case SRF : Achieve x(k) with Eq. (10);
5 Case CCF : Achieve x(k) with Eq. (15);
6 $\alpha k = 1+\sqrt{4\alpha 2 / k-1+1 / 2}$;
7 $z(k) = x(k) + \alpha k-1-1 / \alpha k /(x(k) - x(k-1))$;
8 k = k + 1;
9 until CONVERGENCE;

5. Limitation

Our work is limited in several aspects. First, we assume each name corresponds to a unique single person. Duplicate name can be a practical issue in real-life scenarios. We can extend our method to address this practical problem. For example,

we can learn the similarity between two different names so as to determine how likely the two different names belong to the same person. Second, we assume the top retrieved web facial images are related to a query human name. This is clearly rue for celebrities. However, when the query facial image is not a well-known person, there may not exist many relevant facial images on the WWW. This is a common limitation ofall existing data-driven annotation techniques. This might be artially solved by exploiting social contextual information.

6. Conclusion

This paper investigated a promising search-based face annotation framework, in which we focused on tackling the critical problem of enhancing the label quality and proposed an Unsupervised Label Refinement (ULR) algorithm. We also proposed a Clustering-based Approximation (CBA) solution, which successfully accelerated the optimization task without introducing much performance degradation. From an extensive set of experiments, we found that the proposed technique achieved promising results under a variety of settings. Future work will address the issues of duplicate human names and explore supervised/semi-supervised learning techniques to further enhance the label quality with affordable human manual refinement efforts.

References

[1] S. C. H. Hoi, J. Luo, S. Boll, D. Xu, and R. Jin, Eds., Social Media Modeling and Computing. Springer, 2011. 1
[2] S. Satoh, Y. Nakamura, and T. Kanade, “Name-it: Naming and detecting faces in news videos,” IEEE MultiMedia, vol. 6, pp. 22–35, 1999. 1
[3] P. T. Pham, T. Tuytelaars, and M.-F. Moens, Naming people in news videos with label propagation,” Multimedia, IEEE, vol. 18, no. 3, pp. 44 –55, march 2011. 1
[4] L. Zhang, L. Chen, M. Li, and H. Zhang, “Automated annotation of human faces in family albums,” in ACM Multimedia, 2003. 1
[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.” In IEEE CVPR, 2004, pp. 848–854. 1, 2, 3, 8
[6] J. Yang and A. G. Hauptmann, “Naming every individual in news video monologues,” in ACM Multimedia, 2004, pp. 580–587. 1
[7] J. Zhu, S. C. H. Hoi, and M. R. Lyu, “Face annotation using transductive kernel fisher discriminant,” IEEE Transactions on Multimedia, vol. 10, no. 1, pp. 86–96, 2008. 1, 2
[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. PAMI, vol. 22, no. 12, pp. 1349–1380, 2000. 1
[9] S. C. H. Hoi, R. Jin, J. Zhu, and M. R. Lyu, “Semi-supervised svm batch mode active learning with applications to image retrieval,” ACM TOIS, vol. 27, pp. 1–29, 2009. 1
[10] X.-J. Wang, L. Zhang, F. Jing, and W.-Y. Ma, “Annosearch: Image autoannotation by search,” in CVPR, 2006, pp. 1483–1490. 1, 2
[11] 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 TIST, vol. 2, no. 2, p. 13, 2011. 1
[12] P. Wu, S. C. H. Hoi, P. Zhao, and Y. He, “Mining social images with distance metric learning for automated image tagging,” in Proceedings of the fourth ACM international conference on Web search and data mining, ser. WSDM ’11, Hong Kong, China, 2011, pp. 197–206. 1
[13] D. Wang, S. C. H. Hoi, and Y. He, “Mining weakly labeled web facial images for search-based face annotation,” in SIGIR, 2011. 2, 3, 8
[14] P. Belhumeur, J. Hespanha, and D. Kriegman, “Eigenfaces vs. fisherfaces: recognition using class specific linear projection,” IEEE TPAMI, vol. 19, no. 7, pp. 711 –720, 1997. 2

Author Profile

Mr. Dipak Tayade received the B.E. degree in Computer Science Engineering from SSBTs COET Jalgaon in 20011. Now Appeared for M.E. From Godavari College Of Engineering & Technology, Jalgaon, MH.

Mr. Nilesh Y Chaudhari received the M.E. degree in Computer Science Engineering from SSBTs COET Jalgaon. Now he is working with Godavari College Of Engineering & Technology, Jalgaon, MH.