Person-Specific Face Tracking with Online Recognition

Zhaowei Cai[1], Longyin Wen[1], Dong Cao[1], Zhen Lei[1], Dong Yi[1], and Stan Z. Li[1,2,*]
[1]Center for Biometrics and Security Research & National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
[2]China Research and Development Center for Internet of Thing
{zwcai,lywen,dcao,zlei,dyi,szli}@cbsr.ia.ac.cn

***Abstract*— Person-specific face tracking is a challenging task for the trackers which only focus on the appearance of the target face, because distraction always happens and the identity is difficult to maintain. In this paper, we design a framework combining an off-line detector, an on-line tracker and an on-line recognizer to complete the tracking of person-specific face. Recognizer is the key component in our framework, because the most confident target face will be selected by the recognizer from the pool of detected and tracked faces. Since there is no much prior information about the identities available and the face poses change frequently in surveillance scenarios, accurate recognition is extremely difficult and an on-line formulation is required. In order to ensure the precision of identity recognition with different poses, we project the extracted features of faces to a latent space with the help of Canonical Correlation Analysis (CCA) technique, and then these projected features are incrementally trained using an on-line SVM (LASVM). Experimental results demonstrate that our person-specific face tracking outperforms several state-of-the-art face trackers.**

I. INTRODUCTION

Face tracking is an important research topic in the computer vision community, because many applications need to track human face, including video surveillance, human-computer interaction, robotics, etc. However, face tracking still faces numerous challenges, such as illumination, pose variation, occlusion, appearance-similar faces, out-of-view, identity maintaining, and so on. Person-specific face tracking is a challenging subject of face tracking, in which a face with specific identity is required to be tracked in a long term sequence. With the help of person-specific face tracking, the behavior of the specific person can be analyzed continuously.

For general face tracking, complex appearance model is preferred, such as multi-view face tracking [1], [20]. This kind of methods are highly effective for single face tracking in non-cluttered environment, because they represent the face effectively even when large pose variations happen. However, they will fail to track the target face if the target is severely occluded or out of view. In order to re-find the target face, some trackers incorporate detector [4], [6], [8], [9], [16]. Although the detector is useful in long-term target tracking, the robustness of the tracker will be challenged by other appearance-similar objects all the time.

In this paper, we are going to introduce a stable supervised person-specific face tracking method in unconstrained environment, in which an off-line detector, an on-line tracker and an on-line recognizer are efficiently combined. In details, we extract Haar-like feature [19] of target face and adopt Boosting classifier to construct our tracker. A well trained face detector is incorporated to prevent our framework from being impaired by bad tracking outputs and to re-find the target face when the target reappears. To circumvent the distraction of other appearance-similar faces, an effective discriminative recognition module based on incremental learning SVM classifier (LASVM) [3] is required here. We try to incorporate Canonical Correlation Analysis (CCA) [7] into the recognition module to enhance accuracy of the recognizer in unconstrained environment, because CCA has been proven effective in pose-invariant face recognition task [12]. Actually, the person-specific face tracking problem is viewed as an one-to-many recognition problem and our on-line recognizer protects our tracker from recognizing other similar faces as the target one. At the recognizer updating step therefore, the appearance-similar faces except the target in the scene will be trained as negative samples.

The main contributions of this work are summarized as:

- We propose a novel framework of detection-tracking-recognition to track the person-specific face.
- An on-line learning recognition is proposed to maintain the identity of the target face based on an incrementally learned SVM.
- We firstly introduce Canonical Correlation Analysis into person-specific face tracking because of its robustness in pose-invariant face recognition.

II. RELATED WORKS

Numerous face tracking methods focus on face appearance and they are developed based on appearance model. Stern et al. [17] developed an algorithm to adaptively change the color space models throughout the tracking sequences. This methodology can be used to find the optimal combination of color space model and color distribution model, which is robust in tracking faces under varying illumination. In [15], Lui et al. combined a local linearity of an appearance manifold with a new criterion to select a tangent plane for appearance updating in face tracking. This face tracker obtains good performance even when tracking faces undergo large appearance changes. In order to incorporate more face prior into tracker, Wang et al. [20] combined an off-line trained generic face model and an online-learned specific face appearance model in a dynamic Bayesian network.

*Corresponding author

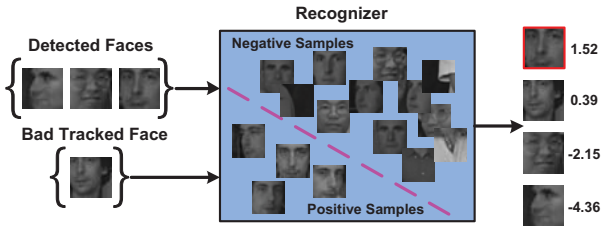

Fig. 1. The function illustration of our recognizer. The recognizer selects the most confident face as the final result from the pool of detected faces and tracked face, hence the bad tracked face will be corrected. The score is the confidence of every face given by our recognizer, and the face with red bounding box is the selected final result.

Thanks to the multiview pose estimation and the robustness of probabilistic principle component analysis to appearance variations, Wang's method tracks faces with pose changes well. Although these appearance based face tracking methods are robust in single face tracking, they cannot recover after the target is totally lost, and they easily skip to other similar faces which are near to the target face.

In order to re-find the target after occlusion or out of view, detector is incorporated. Verma et al. [18] integrated the information of face probabilities provided by the detector and the temporal information provided by the tracker to obtain more stable detection and tracking results. Grabner et al. [6] explored the combination of a fixed detector trained with labeled prior and an on-line classifier updated in semi-supervised way. Similarly, tracking, learning and detection are also combined in [8], [9], which perform very well in unconstrained environments. Unfortunately, these methods are easily distracted by other objects that have similar appearance with the target. Dinh et al. [4] incorporated the concept of supporter and distracter to circumvent distraction problem. Although it handles the skipping problem to some degree, this limit cannot be overcome thoroughly. A work similar to our framework is introduced in [16], in which Stalder et al. combined an off-line detector, supervised on-line recognizer and a semi-supervised on-line tracker together to track the target. The recognizer is updated conservatively only when the samples are validated by the detector, but the recognizer does not always have accurate performance.

III. Supervised Face Tracking Framework

The recognizer is the key component in our tracking framework. Its most important function is to effectively combine the detection and tracking results, that is to decide which face state is the most confident one among the set of detected faces and tracked face. When the tracking outputs are less confident, for example, the tracker totally loses the target or drifting away happens, the tracked face x_T provided by the tracker is un-trustable. Instead, the detected faces set $\{x_D^i\}_1^n$ provided by the detector is more useful to re-find the initial target face, because at least the detected objects are confident faces and less background will be included. After performing recognizer on $\{x_D^i\}_1^n$ and x_T, the most confident candidate will be recognized as the target face x^* if its confidence is larger than a conservative re-find threshold θ_r. The updating step will be activated if the confidence of the target face is larger than another conservative updating threshold θ_u. The function of recognizer is shown in Fig. 1 and the algorithm details are described in Algorithm 1.

Algorithm 1 Person-Specific Face Tracking

Initialize Boosting+Haar feature based tracker.
Initialize LASVM+CCA feature based recognizer.
while *run* **do**
  Perform face detector to get detected face set $\{x_D^i\}_1^n$.
  Perform tracker to get tracked face x_T.
  Perform recognizer on detected face set $\{x_D^i\}_1^n$ and get the confidence set $\{conf_D^i\}_1^n$. Find the face state x_D^{max} with maximal confidence $conf_D^{max}$.
  Perform recognizer on tracked face x_T and get its confidence $conf_T$.
  if $conf_T < \theta_r$ and $conf_D^{max} > \theta_r$ **then**
    Accept x_D^{max} as the final target face x^*, and its confidence $conf_{x^*} = conf_D^{max}$.
  else
    Accept x_T as the final target face x^*, and its confidence $conf_{x^*} = conf_T$.
  end if
  if $conf_{x^*} > \theta_u$ **then**
    Update Boosting tracker with x^*.
    Update LASVM recognizer. x^* is trained as positive sample, and the other faces in $\{x_D^i\}_1^n$ which has no overlap with x^* and some randomly selected background patches are trained as negative samples.
  end if
end while

For detection, we prefer to employ AdaBoost as our detection classifier [19] and Multi-block Local Binary Patterns (MB-LBP) [13] as feature here because of the robust and efficient performance of their combination [22]. The CCA based recognition needs to resize the cropped faces into uniform form, and we simply transform the faces by aligning two eyes to two fixed coordinates with a similarity transformation. Specifically, we adopt the eye detection in [21] which is designed for surveillance scenarios.

IV. Online Face Tracker

The face tracking problem is formulated as a state estimation and the motion model is viewed as a Markovian state transition process. Let $Z_{1:t} = \{z_1, \cdots, z_t\}$ represent the observation data up to time t. $x_t = (l_t, b_t)$ is the state of the target at time t, which contains the position l_t and size b_t of the target. In our tracker, l_t is composed by X-coordinate and Y-coordinate, and b_t consists of width and height. The posterior probability is estimated as the recursive equation:

$$p(x_t|Z_{1:t}) \propto p(z_t|x_t) \int p(x_t|x_{t-1}) p(x_{t-1}|Z_{1:t-1}) dx_{t-1} \quad (1)$$

where $p(z_t|x_t)$ is the likelihood of the candidate sample, $p(x_t|x_{t-1})$ is the state transition probability and

$p(x_{t-1}|Z_{1:t-1})$ is the state estimation probability given all observations up to time $t-1$. For state transition, we assume l_t is independent of b_t, so

$$\begin{aligned} p(x_t|x_{t-1}) &= p(l_t, b_t|l_{t-1}, b_{t-1}) \\ &= p(l_t|l_{t-1})p(b_t|b_{t-1}) \end{aligned} \tag{2}$$

As suggested in [2], the position transition term is directly specified as $p(l_t|l_{t-1}) = 1$, if $\|l_t - l_{t-1}\| < r$, otherwise $p(l_t|l_{t-1}) = 0$, where r is the search radius. The scale transition term $p(b_t|b_{t-1})$ is similarly handled as [2].

In this paper, we view face tracking as a binary classification between the face and the background, because face appearance has obvious difference from the background. The works of Viola et al. [19] and Grabner et al. [5] make the online rapid face tracking based on Boosting+Haar available, and we also choose this strategy. The candidate sample likelihood $p(z_t|x_t)$ is computed as:

$$p(z_t|x_t) \propto \exp\{\sum_{i=1}^{N} h_i\} \tag{3}$$

where h_i is the selected weak classifier corresponding to the Haar-like feature. By linearly combining the selected N weak classifiers, a strong classifier is constructed and the confidence map can be obtained. Then, the state of optimal tracked face can be estimated by MAP solution:

$$x_T = \arg\max_{x_t} p(x_t|Z_t) \tag{4}$$

V. On-line Face Recognizer

Generally speaking, single face tracking [18], [9], [15], [17], [20] does not require recognition, multiple faces tracking [14] demands accurate many-to-many recognition and video-based face recognition [10], [11] integrates tracking for better recognition. Differently, what the person-specific face tracking needs is one-to-many recognition. Some subspace based methods, LDA [23] and CCA [12] for instance, have been proven effective in one-to-many face recognition. LDA strongly assumes the covariance matrix is identical for different classes and the data distribution density is Gaussian. It is clear that the face data distribution with different pose variations does not subject to Gaussian distribution. Therefore, LDA technique is not perfectly applicable here. Instead, CCA technique that carefully learns the relationship between different distributions and constructs the most correlated latent space meets the requirements in surveillance scenarios. With the help of CCA technique, the feature distance between the faces with different poses of same identity is closer in latent space than that in original feature space. And thus, the recognition accuracy will be enhanced, which consequently improves the tracking robustness of our framework. After feature extraction by CCA technique therefore, we introduce an incremental binary SVM classifier to complete one-to-many recognition task, which will robustly adapt to the appearance variations of the target face during the tracking process.

A. Canonical Correlation Analysis based Latent Space

CCA [7] is a technique that learns a set of M different projectors from a set of observed content under M different styles. The projections of different styles of a particular content are maximally correlated in the projected space. Therefore, it can be used in pose-invariant face recognition, in which the projections of different pose faces will be highly correlated in the latent space. Given the training data pairs $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]$, $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_n]$. The leading factor subspaces are the linear subspaces of the training data sets $\mathbf{X}$ and $\mathbf{Y}$, with a retained dimensionality d. CCA takes into account the two data sets simultaneously and finds the optimal linear projective matrices, also called canonical projection pairs, $\mathbf{W}^x = [\mathbf{w}_1^x, \cdots, \mathbf{w}_d^x]$ and $\mathbf{W}^y = [\mathbf{w}_1^y, \cdots, \mathbf{w}_d^y]$, from the corresponding data $\{\mathbf{X}, \mathbf{Y}\}$, such that $\mathbf{x}_i' = \mathbf{X}^T\mathbf{w}_i^x$ and $\mathbf{y}_i' = \mathbf{Y}^T\mathbf{w}_i^y$ are most correlated. This is done by maximizing the following correlation

$$\begin{aligned} \rho(\mathbf{w}_i^x, \mathbf{w}_i^y) &= \frac{E[\mathbf{x}_i'^T\mathbf{y}_i']}{\sqrt{E[||\mathbf{x}_i'||^2]E[||\mathbf{y}_i'||^2]}} \\ &= \frac{\mathbf{w}_i^{xT}\mathbf{C}_{xy}\mathbf{w}_i^y}{\sqrt{\mathbf{w}_i^{xT}\mathbf{C}_{xx}\mathbf{w}_i^x\mathbf{w}_i^{yT}\mathbf{C}_{yy}\mathbf{w}_i^y}} \\ s.t. \quad \rho(\mathbf{w}_j^x, \mathbf{w}_i^y) &= \rho(\mathbf{w}_i^x, \mathbf{w}_j^y) = 0 \\ for \quad j &= 1, \cdots, i-1 \end{aligned} \tag{5}$$

where $\mathbf{C}_{xy}$, $\mathbf{C}_{xx}$ and $\mathbf{C}_{yy}$ are the correlation matrices computed from the training data sets $\mathbf{X}$ and $\mathbf{Y}$. Thus, the solution $\mathbf{W}^x$ and $\mathbf{W}^y$ can then be obtained by solving the generalized eigenproblem:

$$\mathbf{AW} = \mathbf{BW}\Lambda \tag{6}$$

where,

$$\mathbf{A} = \begin{pmatrix} 0 & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \mathbf{C}_{xx} & 0 \\ 0 & \mathbf{C}_{yy} \end{pmatrix} \tag{7}$$

After performing CCA on two data sets, we can extract the most correlative component pairs from the original data. The general optimal linear projection matrices are designed between frontal faces and profile faces with specific pose. In tracking sequences however, face poses may vary from -90° to 90°, and sometimes even larger degrees. An intuitive solution is to design multiple projection matrices specified for multiple poses training samples respectively, but we decide not to adopt this method because the errors of face pose estimation will also undermine the accuracy of recognition. Instead, we project profile face images in different poses with $\mathbf{W}^y$ into a common latent space which is closely related with the latent space projected by $\mathbf{W}^x$. Therefore, during the process of training CCA model, we collect training profile face samples with -30°, -15°, 15° and 30° poses into $\mathbf{Y}$, and the same number of frontal face samples into $\mathbf{X}$ corresponding to $\mathbf{Y}$. Then, the correlative projection subspaces $\mathbf{W}^x$ and $\mathbf{W}^y$ for multiple poses can be obtained. The complimentary information between adjacent pose projections will improve the robustness of pose-invariant face recognition. Although the faces with different poses of same subject may be not

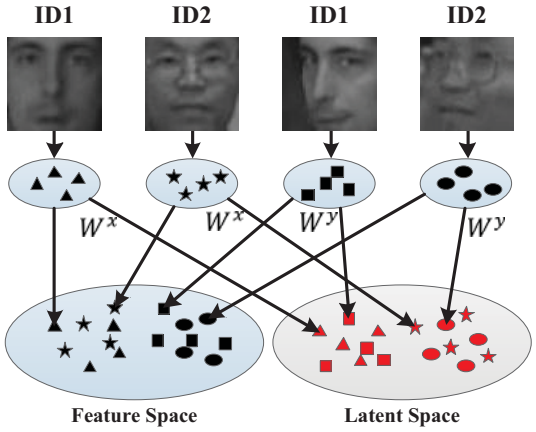

Fig. 2. The illustration of feature space and latent space projected by CCA technique. The feature distance between the faces with different poses of same identity is closer in latent space than that in original feature space. Therefore, pose-invariant face recognition is available in the latent space.

recognized as the same identity in the original feature space, their feature distance will be closer after the features being projected into CCA latent space, as illustrated in Fig. 2

B. Online SVM Training

Since the training samples will be collected on-the-fly in surveillance scenarios, an online recognition method should be developed to meet the demand, and an online SVM, which is called LASVM [3], is adopted. Given the training samples, the linear discriminant function can be written as a kernel expansion in SVM framework:

$$\widehat{y}(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i) + b \tag{8}$$

The coefficients α_i in above equation are found by defining the dual objective function with training sample (x_i, y_i):

$$W(\alpha) = \sum_i \alpha_i y_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) \tag{9}$$

and the support vectors can be obtained by solving the SVM Quadratic Programming problem. The LASVM algorithm contains two alternating direction search procedures, PROCESS and REPROCESS, to update the support vectors. Please refer to [3] for more details.

The recognition classifier is initialized with the manually labeled sample in the first frame and updated online. The positive samples are the successfully tracked faces with the recognition confidence larger than a conservative threshold θ_u. At the same time, the negative samples are the other faces in the tracking scene except the target face. In order to collect more negative samples for better performance of LASVM, we also randomly collect 5 image patches in the background as negative samples at every sample collection step. The feature $F = \{f_1, \cdots, f_d\}$ of a sample is the projected vector of reshaped image pixel values by CCA projection matrix $\mathbf{W}^x$ or $\mathbf{W}^y$. The number of retained dimensionality during the process of eigenvalue decomposition is $d = 50$, and LASVM is updated every three frames in our experiment.

A simple and intuitive solution is adopted here to decide which one of $\mathbf{W}^x$ and $\mathbf{W}^y$ should be used to project the original pixel feature. At first, we obtain two projected vectors by both $\mathbf{W}^x$ and $\mathbf{W}^y$, and they are simultaneously scored by recognizer with Equ. 8. For positive sample, the projected vector with higher score must be closer to the positive data distribution in training pool and will be viewed as positive training sample. For negative sample, the projected vector with lower score must be closer to the negative data distribution in training pool and will be viewed as negative training sample. The properties of online recognizer and CCA latent space make it reasonable, and the satisfactory tracking performance also demonstrates the soundness of our framework.

VI. Experiments

In order to evaluate the performance of our tracker, we overall track 4 person-specific faces on 3 challenging video sequences (*multiface*, *babies* and *juli*). These sequences include most of the challenges: large pose variations, scaling, appearance-similar faces, wholly occlusion, out of view, etc. Our tracker is implemented in C++ code and runs approximately 4-8 fps on the standard PC platform with 2.4 GHz CPU and 3 GB memory. In our experiment, the re-find threshold $\theta_r \in [0, 0.5]$ and update threshold $\theta_u \in [0.5, 1]$, and other parameters are fixed.

For the feature vector projected by CCA, the pixel value vector of a reshaped 60×60 gray image patch is reduced into a 50 dimension feature vector, and we call it CCA feature later for simplicity. To demonstrate the superiority of CCA based recognition, we replace the CCA with Histogram of Oriented Gradients (HOG) and Linear Discriminative Analysis (LDA). The HOG feature vector are created from resized 24×24 gray image patch, in which block size is 2×2, cell size is 8×8, strike size is 4 and a 9-bins oriented histogram is adopted, and the HOG vector size is 324. For the feature vector projected by LDA, the pixel value vector of a reshaped 32×32 gray image patch is reduced into a 231 dimension feature vector in latent space by LDA technique, and we will call it LDA feature similarly. The feature dimensions of CCA, HOG and LDA are heuristic setting so that they acquire relatively better performance and higher efficiency. The framework with these three different recognition modules are called as CCA, LDA and HOG respectively.

For fair comparison, the trackers with detector, TLD [8], ContextT [4] and SemiBoost [6], are chosen to complete the quantitative experiment, because they own advantages to recover from wholly occlusion and face disadvantages to be distracted by other similar objects, similar to our framework. BeyondBoost [16] is also compared because it similarly introduces recognizer to improve tracking performance. We utilize the default parameters of these trackers which are provided in their papers or codes. We choose the best one of 5 runs, and for each run, we only change the search radius.

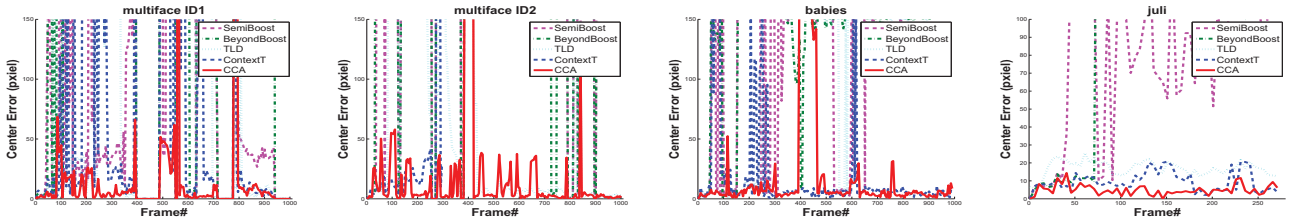

Fig. 3. Tracking results of our tracker (CCA), SemiBoost, BeyondBoost, TLD and ContextT.

Seq.	SemiBoost	BeyondBoost	TLD	ContextT	HOG	LDA	CCA
multiface(ID1)	128	327	144	71.2	123	82.5	**15.0**
multiface(ID2)	346	341	89.7	35.0	49.3	60.4	**22.4**
babies	171	156	129	18.4	132	30.5	**17.0**
juli	85.5	136	16.9	9.64	**4.42**	7.20	4.98

TABLE I

THE COMPARISON OF AVERAGE ERROR CENTER LOCATION IN PIXEL.

Seq.	SemiBoost	BeyondBoost	TLD	ContextT	HOG	LDA	CCA
multiface(ID1)	339	334	583	703	563	633	**828**
multiface(ID2)	135	150	683	234	768	698	**798**
babies	245	120	210	559	235	744	**889**
juli	45	60	204	154	**274**	259	**274**

TABLE II

THE COMPARISON OF SUCCESSFULLY TRACKED FRAMES.

A special note is that the results of ContextT are obtained by running the executable code provided by Dinh, which are slightly different from the reported results in [4]. All of the experimental results are presented in Table I, Table II, Fig. 3 and Fig. 4.

multiface This 1002 frames sequence is captured indoor, containing most of challenges for person-specific face tracking, such as large pose variation, wholly occlusion, distracting face, out-of-view, etc. We do not track all of the faces in the sequence but choose ID1 and ID2 because they are more challenging than the other two people. The similarity between the faces is really a great challenge for most of trackers based on detector, thus TLD, SemiBoost, BeyondBoost and ContextT frequently shift to other faces. Because of the robustness of CCA in pose-invariant face recognition, our tracker nearly does not shift to other faces. Meanwhile, the face detection results will also help to correct the updating samples and prevent the tracker from being drifted away. Consequently, our method performs well when the pose of target face varies dramatically, and can re-find the target precisely even when the tracker does not have accurate outputs for a while. In this sequence, ID1 is being out-of-view from frame 395 to frame 490 and from frame 715 to frame 775. It is very difficult to recognize ID1 before and after out-of-view as the same identity for most trackers. TLD, SemiBoost, BeyondBoost and ContextT cannot track ID1 in time after it reappears without an efficient recognizer. Differently, our method quickly re-finds ID1 after out-of-view and ID2 after wholly occlusion, as shown in Fig. 4.

babies This 994 frames sequence is also captured indoor. Three triplet babies who have extremely similar faces are on the floor, playing themselves. Certainly, TLD, SemiBoost and BeyondBoost frequently shift to the other two faces and they almost cannot recover the target face. Furthermore, ContextT which incorporates the supporters and distracters still frequently loses the target. Although our tracker is also distracted by one triplet from frame 390 to frame 459, as shown in Fig. 3, it recovers and outputs good tracking results after the distraction. Somtimes, the background patches will be recognized as target by SemiBoost, BeyondBoost and ContexT, as displayed in Fig. 4. With the help of the face detector and the robust boosting based tracker, our tracker nearly will not recognize the background as the target and tracking the face very well when large variations happen.

juli This 274 frames sequence is our own which is cut from the film "Flipped". Juli in this video clip undergoes large scaling and illumination variations. Since SemiBoost and BeyondBoost do not consider scaling, they lose the target face quickly and almost cannot recover. As shown in Table I, Table II and Fig. 4, TLD and ContextT perform very well, while our framework still outputs the best tracking results, which demonstrates the robustness of our framework to illumination and scaling.

Overall speaking, CCA owns higher recognition accuracy over HOG and LDA in our experiments by considering the correlation between frontal and profile faces. LDA neglects the data distribution difference between different poses of the same identity, so its performance is a little worse. HOG works in the original feature space and definitely no correlation between different poses is constructed. In the sequence *juli*, HOG, LDA and CCA have comparative results and HOG performs the best, because the pose of target face is nearly not changed in this sequence and original features perform well. In the other two sequences however, the advantages of CCA are obvious, which can be demonstrated in Table I and Table II.

VII. CONCLUSIONS

A novel person-specific face tracking method is proposed in this paper, in which an off-line face detector, an online tracker and an online recognizer are efficiently combined. Boosting is applied as the classifier in detector and tracker, and the features are MB-LBP and Haar respectively. Considering the good performance of Canonical Correlation Analysis in pose-invariant face recognition, we incorporate it into our online recognizer, combined with an online classifier

Fig. 4. Tracking results. The results of SemiBoost [6], BeyondBoost [16], TLD [8], ContextT [4], and our tracker (CCA) are depicted as magenta, green, cyan, blue and red rectangles respectively.

LASVM. The superior performance in challenging sequences proves the robustness of our framework.

VIII. ACKNOWLEDGMENTS

This work was supported by the Chinese National Natural Science Foundation Project #61070146, #61105023, #61103156, #61105037, #61203267, National IoT R&D Project #2150510, National Science and Technology Support Program Project #2013BAK02B01, Chinese Academy of Sciences Project No. KGZD-EW-102-2, European Union FP7 Project #257289 (TABULA RASA http://www.tabularasa-euproject.org), and AuthenMetric R&D Funds.

REFERENCES

[1] K. H. An, D. H. Yoo, S.-U. Jung, and M. J. Chung. Robust multi-view face tracking. In *IROS*, pages 1905–1910, 2005.
[2] B. Babenko, M.-H. Yang, and S. Belongie. Robust object tracking with online multiple instance learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(8):1619–1632, 2011.
[3] A. Bordes, S. Ertekin, J. Weston, and L. Bottou. Fast kernel classifiers with online and active learning. *Journal of Machine Learning Research*, 6:1579–1619, 2005.
[4] T. B. Dinh, N. Vo, and G. G. Medioni. Context tracker: Exploring supporters and distracters in unconstrained environments. In *CVPR*, pages 1177–1184. IEEE, 2011.
[5] H. Grabner, M. Grabner, and H. Bischof. Real-time tracking via on-line boosting. In *BMVC*, pages 47–56, 2006.
[6] H. Grabner, C. Leistner, and H. Bischof. Semi-supervised on-line boosting for robust tracking. In *ECCV (1)*, pages 234–247, 2008.
[7] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3-4):321–377, 1936.
[8] Z. Kalal, J. Matas, and K. Mikolajczyk. P-n learning: Bootstrapping binary classifiers by structural constraints. In *CVPR*, pages 49–56, 2010.
[9] Z. Kalal, K. Mikolajczyk, and J. Matas. Face-tld: Tracking-learning-detection applied to faces. In *ICIP*, pages 3789–3792, 2010.
[10] M. Kim, S. Kumar, V. Pavlovic, and H. A. Rowley. Face tracking and recognition with visual constraints in real-world videos. In *CVPR*, 2008.
[11] K.-C. Lee, J. Ho, M.-H. Yang, and D. J. Kriegman. Visual tracking and recognition using probabilistic appearance manifolds. *Computer Vision and Image Understanding*, 99(3):303–331, 2005.
[12] A. Li, S. Shan, X. Chen, and W. Gao. Maximizing intra-individual correlations for face recognition across pose differences. In *CVPR*, pages 605–611, 2009.
[13] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li. Learning multi-scale block local binary patterns for face recognition. In *ICB*, pages 828–837, 2007.
[14] R. Liu, X. Gao, R. Chu, X. Zhu, and S. Z. Li. Tracking and recognition of multiple faces at distances. In *ICB*, pages 513–522, 2007.
[15] Y. M. Lui, J. R. Beveridge, and L. D. Whitley. Adaptive appearance model and condensation algorithm for robust face tracking. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 40(3):437–448, 2010.
[16] S. Stalder, H. Grabner, and L. V. Gool. Beyond semi-supervised tracking: Tracking should be as simple as detection, but not simpler than recognition. In *OLCV*. IEEE, September 2009.
[17] H. Stern and B. Efros. Adaptive color space switching for face tracking in multi-colored lighting environments. In *FGR*, pages 249–254, 2002.
[18] R. C. Verma, C. Schmid, and K. Mikolajczyk. Face detection and tracking in a video by propagating detection probabilities. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(10):1215–1228, 2003.
[19] P. A. Viola and M. J. Jones. Rapid object detection using a boosted cascade of simple features. In *CVPR (1)*, pages 511–518, 2001.
[20] P. Wang and Q. Ji. Robust face tracking via collaboration of generic and specific models. *IEEE Transactions on Image Processing*, 17(7):1189–1199, 2008.
[21] D. Yi, Z. Lei, and S. Z. Li. A robust eye localization method for low quality face images. In *IJCB*, pages 1–6, 2011.
[22] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li. Face detection based on multi-block lbp representation. In *ICB*, pages 11–18, 2007.
[23] W.-Y. Zhao, R. Chellappa, and A. Krishnaswamy. Discriminant analysis of principal components for face recognition. In *FG*, pages 336–341, 1998.