INTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING
DOI: 10.46300/9106.2021.15.76 Volume 15, 2021

Face Tracking via Content Aware Correlation Filter

Houjie Li, Shuangshuang Yin, Fuming Sun, Fasheng Wang*

Dalian Minzu University,
No. 18, West Liaohe Road, Jinpu New District, Dalian 116600
China

Received: February 10, 2021. Revised: June 26, 2021. Accepted: July 19, 2021. Published: July 20, 2021.

Abstract- Face tracking is an importance task in many computer vision based augment reality systems. Correlation filters (CFs) have been applied with great success to several computer vision problems including object detection, classification and tracking, but few CF-based methods are proposed for face tracking. As an essential research direction in computer vision, face tracking is very important in many human-computer applications. In this paper, we present a content aware CF for face tracking. In our work, face content refers to the locality sensitive histogram based foreground feature and the learning samples extracted from complex background. It means that both foreground and background information are considered in constructing the face tracker. The foreground feature is introduced into the objective function which could learn an efficient model to adapt to the face appearance variation. For evaluating the proposed face tracker, we build a dataset which contains 97 video sequences covering the 11 challenging attributes of face tracking. Extensive experiments are conducted on the dataset and the results demonstrate that the proposed face tracker shows superior performance to several state-of-the-art tracking algorithms.

Keywords- Face tracking, Correlation Filters, Locality Sensitive Histogram

I. Introduction

WITH the repaid development of computer vision and virtual reality, augmented reality systems are becoming more common in many real-world applications, especially in healthcare [1, 2]. Augmented reality is one of the most promising digital health technologies at present. Many of the existing healthcare application systems require accurate and real-time tracking of human faces in order to finish specific tasks [3, 4, 5]. Face tracking is a fundamental task in related intelligent human-computer interaction systems. This is because face tracking is an essential step to build 3D face models [6, 7]. A good face tracking method should track human face accurately in a variety of lighting conditions, head poses, environments, and occlusions [8].

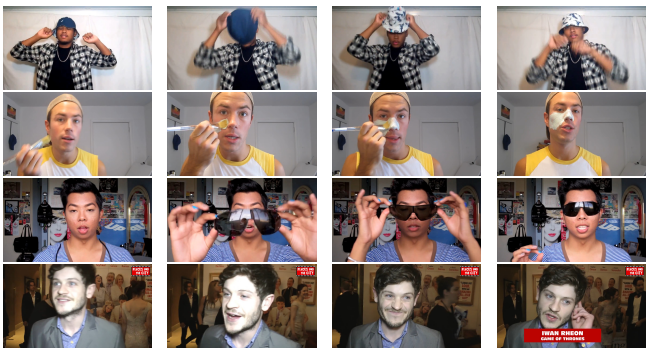

Fig. 1: Challenges in face tracking.

As is known to all, the main challenges in general object tracking include illumination variation (IV), fast motion (FM), in-plane rotation (IPR), scale variation (SV), motion blur (MB), out-of-plane rotation (OPR), occlusion (OCC), deformation (DEF), background clutters (BC), out-of-view (OV) and low resolution (LR) [9]. Face tracking also suffers these challenges. But different from general object tracking, some of the main challenges in face tracking may stem from the appearance variations induced by heavy makeup, face decorations, severe facial expression changes, and so on (see Fig. 1), which makes face tracking a challenging task in various real-world applications.

In the past two decades, face tracking has been intensively studied in the literature and many advanced algorithms are proposed for face tracking. The existing tracking algorithms can be classified using different standards, one of which is based on whether the whole face is tracked or individual facial features are tracked (facial landmark tracking). There exists a lot of facial landmark tracking methods in the literature [10, 11, 12, 13, 14]. When the whole face is considered during tracking, it can be represented as a general object. Different from

[0]This work was supported by National Natural Science Foundation of China (Grant No. 61972068, 61976042, 62072152), Liaoning Baiqianwan Talent Program (Grant No. 2018B09), LiaoNing Revitalization Talents Program (Grant No. XLYC2007023), Innovative Talents Program for Liaoning Universities (Grant No. LR2019020), and Liaoning Natural Science Foundation (Grant No. 2019-ZD-0171, 20180550625, 2019-ZD-0182).
Corresponding: wangfasheng@dlnu.edu.cn

facial landmark tracking, this kind of face tracking aims to localize the exact position of the whole face or head.

Most of the existing face trackers adopt hand-crafted features, such as skin color [15] and geometry [16, 17], Harr-like feature [18], optical flow [19], and so on. Traditional tracking methods like mean shift [20], multiple instances and Online Adaboost [21], particle filters [22], Kalman filter [12, 18], bilateral filtering [23] have been applied successfully in face tracking. Recently, as deep learning has gained special attention in computer vision field, many deep learning based face trackers are proposed in the literature [24, 25, 26, 27, 28]. The main drawback of deep trackers lie in that researchers need strong-computation power to perform online/offline model learning which is quite time-consuming. Another popular tracking framework, correlation filters, also have been applied with success to face verification [29], face recognition [30, 31] and face tracking [32, 33, 34]. Since its first application in object tracking [35], a lot of CF-based trackers are proposed which show improved performance in object tracking [36, 37, 38, 39, 40, 41]. However, the existing CF-based trackers suffer from boundary effects due to the circulant shifted sampling process. Another problem is that the training samples are obtained by shifting the original base sample cirularly, which means that the generated samples are virtual samples. At the same time, the samples are generated from foreground target object lacking negative samples from background area which deteriorates the tracking performance in complex and cluttered background.

In order to evaluate the performance of face trackers, a number of video sequences must be collected to run the trackers and compare their performances using specific evaluation metrics. Shen et. al. [11] developed a facial landmark tracking dataset, 300-VW, which contains 110 video sequences. Chrysos et. al. [10] perform comprehensive evaluation of facial landmark tracking method on 300-VW dataset. But this dataset is not suitable for face tracking as it focuses much on facial landmark detection and tracking, while the common challenges mentioned above are not included in this dataset. In [42], Lin et. al. proposed a mobile face tracking which consists of 80 mobile videos. MobiFace dataset contains most of the challenging attributes in online tracking benchmark (OTB), but it does not contain the challenges induced by heavy makeups, face decorations, severe facial expression changes, etc. On the other hand, the sequences are captured by mobile phones which are not suitable for general performance evaluation purpose.

In this paper, we design a CF-based face tracking method and collect a face tracking dataset for performance evaluation. The proposed face tracker considers not only the foreground face feature but also background information of faces. First, negative samples extracted from the background are used for learning CFs. Second, for the foreground feature, we compute the locality sensitive histogram (LSH) [43] based feature of the face and incorporate the feature into the CF model, which can enhance the discriminative ability of the face tracker. The CF model is solved using the alternating direction method of multipliers (ADMM) [44]. In addition, we collect a face tracking dataset which contains both indoor and outdoor situations and includes 11 attributes indicating different challenges in face tracking. We conduct extensive experiments on the collected dataset. Experimental results show that our proposed method outperforms several state-of-the-art tracking methods.

II. Related Works

A. Correlation Filter Tracking

In the past ten years, CF-based tracking methods has gained special attention in object tracking. Bolme et. al. [35] first proposed a CF tracker (MOSSE) for object tracking with a very fast speed of about 700 frames per second. In MOSSE tracker, the filter is trained with only grayscale samples which limits the application of MOSSE in other challenging scenarios. Many improved CF-based trackers learn multi-channel filters on HOG feature or color names feature [36, 38, 40, 45]. Many works on face tracking focus on correlation filters. In [32], My et. al. combine an adaptive CF and Viola-Jones face detection method to design a robust real-time face tracking algorithm for mobile robot. The designed algorithm can help the robot track human face as well as the facial features of eye corners and nose under different illumination conditions.

Gaxiola et. al. [33] proposed a locally-adaptive correlation filter for face tracking. A composite correlation filter which is adapted online is used to detect and locate faces in each frame of a video sequence. In [34], Su et al. propose a fast face tracker based on the kernelized correlation filter (KCF) [36]. Multi-task cascade convolutional neural networks (MTCNNs) are used for detecting face. Liao et al. [46] apply the KCF to driver face tracking by combining the MTCNN and deepSORT method. Soldic et al. [47] developed a multi-face tracking system by combining discriminative scale space tracking (DSST [45]) and a robust face detector. The proposed system could handle long-term full occlusions.

B. Deep Learning based Tracking

Deep learning have shown its strong ability in many real-world applications [48]. In the past decades, many deep learning methods, including CNNs, Siamase network and generative adversarial network (GAN), have been applied with great success in object detection, classification, recognition and tracking. In [25], convolutional neural networks (CNNs) are used for face tracking. The authors take advantage of the strong representation ability of hierarchical CNN features. Discriminative face information is captured at both local and global level using two types of Siamese CNNs, Local-CNNs (L-CNNs) and Global-CNNs (G-CNNs). The L-CNNs are used to extract local features from target area, such as eye, nose and mouth. The G-CNNs are designed to extract global features from the entire face. A correlation filter tracking framework is used to integrated the two-level features to construct a robust face tracker.

Li et. al. [26] proposed a simple real-time multi-face tracking system which is composed of three parts: face detection module, feature extraction module and tracking module. The authors adopt multi-task CNN to detect human face, and use a simple CNN to obtain face features of the detected faces. A shallow network is used to track target face on the basis of the extracted features. Lian et. al. [27] designed a real-time face tracking system using multi-task CNN for face detection. The authors aim to solve face occlusions or fast motion which induce tracking failure. Multiple features which include appearance, motion and shape features are fused to enhance the robustness of face tracking. Males et. al. [49] proposed a multi-agent dynamic system which can be easily adapted for robust multi-face tracking problem. Deep learning method are used for face detection which is integrated in the proposed tracking system.

In [28], the authors design a dual-agent deep reinforcement learning algorithm for deformable face tracking. A unified framework are designed which can simultaneously generate bounding box and perform face alignment tasks. The deep reinforcement learning is used to train the dual agent models which is responsible for exploiting the relationships of the two tasks.

C. Other Face Trackers

Zou et. al. [50] proposed to perform face tracking in the gradient logarithm field (GLF) feature space in order to overcome the low-resolution and illumination changes problems. The proposed GLF feature is a global feature which mainly depends on the intrinsic characteristic of a face and is illumination insensitive. In [51], the authors propose to describe a face in a L2-subspace using a relational graph, and proposed a robust face tracking method which specify an importance to appearance features during tracking initialization and the whole face tracking process. They design a weighted score-level fusion scheme to localize target face from output of the tracker that have the highest fusion score. Huang et. al. [21] use multiple instance and online AdaBoost to train a face tracking model and incorporate face detection method to recover tracking when occlusions are detected.

In order to solve the drifting problem encountered in face tracking, in [52], Jiang et. al. employ a supervised descent method (SDM) and a compressive tracking method (CT) to propose a robust face tracking algorithm. The SDM is employed for correcting drifting errors of CT during frontal face tracking. When face orientation changes severely, SDM tracking failure occurs. The authors switch tracking to CT to keep tracking until SDM recover from tracking failure.

Li et. al. [53] designed a face tracker by fusing multiple features within the particle filter framework. Color histogram and edge orientation histogram are used for describing the facial feature while the features are fused using a self-adaptive strategy in order to compute the particle weight. Wu et. al. [54] proposed a coupled hidden Markov random field (CHMRF) for simultaneously modeling face clustering and face tracking. Two HMRF models are used for clustering faces and tracklet linking. The authors provided a method for joint optimization of cluster labels and face tracking. Nam et. al. [55] propose a face tracking system which comprises detection, tracking and false track removal. The multi-view deformable part-based model (DPM) is used as face detector, and a tracking-by-detection framework is leveraged for tracking adding motion cues, color histogram and SURF features. Wrong detections are filtered out using an explicit false alarm removal step.

Aspandi et. al. [56] build a fully end-to end facial tracking model from Re^3 tracker [57]. The proposed model has a long short term memory layer (LSTM) which could model the short and long temporal dependency between frames. The authors perform extensive experiments using 300-VW dataset [11]. Experimental results show that the proposed model perform superior to several advanced face trackers.

III. Proposed Method

A. Correlation Filters

The goal of the standard discriminative correlation filter (DCF) is to learn a multi-channel CF $\mathbf{h}$ in the spatial domain based on training examples $\{(\mathbf{x}_k, \mathbf{y}_k)\}_{k=1}^{t}$, where $\mathbf{x}_k$ is training sample with d channels, and $\mathbf{y}_k$ is the correlation response. The learning process can be formulated as minimizing the objective function as follow:

$$\varepsilon(\mathbf{h}) = \frac{1}{2}\left\|\mathbf{y} - \sum_{k=1}^{K} \mathbf{x}_k * \mathbf{h}_k\right\|_2^2 + \frac{\gamma}{2}\sum_{k=1}^{K}\|\mathbf{h}_k\|_2^2 \qquad (1)$$

where $\mathbf{y} \in R^D$ denotes the desired correlation response, K denotes the number of feature channels, $\mathbf{h}_k$ is the kth channel of the filter, γ is the weight of regularization term, and $*$ denotes the correlation operator. According to [35], (1) can be considered as solving ridge regression problem in the spatial domain using the following objective function:

$$\varepsilon(\mathbf{h}) = \frac{1}{2}\sum_{j=1}^{D}\left\|\mathbf{y}(j) - \sum_{k=1}^{K} \mathbf{h}_k^{\top}\mathbf{x}_k[\triangle\tau_j]\right\|_2^2 + \frac{\gamma}{2}\sum_{k=1}^{K}\|\mathbf{h}_k\|_2^2 \qquad (2)$$

where $\mathbf{y}(j)$ is the jth element of correlation response $\mathbf{y}$, $\triangle\tau_j$ represents the circular shift operator.

As the baseline CF suffers from the annoying boundary effects caused by the circulant shifted samples, a spatial regularization term is introduced into the objective function to penalized the filter coefficient in the learning process [38] to alleviate the boundary effects. But fixed spatial regularization weight cannot adapt the tracker to complex tracking scenarios. Another problem is that all the training samples are generated from circular shifted foreground patches which ignores the background information. Most of the existing CF trackers only use histogram of gradient (HOG) feature which is insufficient for face tracking when encountering severe appearance changes.

B. Locality sensitive Histogram

Locality sensitive histogram is a location-related statistical feature which has been applied with success to visual tracking as it enhances the trackers' ability of separating the target from complex background [58]. It considers every pixel in an image region. Let $H_p^E(b)$ denotes the bin b of LSH computed at pixel location p, where $b = 1, 2, ..., B$. Suppose we have an image I,then $H_p^E(b)$ can be computed as follow:

$$H_p^E(b) = \sum_{q=1}^{N} \gamma^{|p-q|} \cdot Q(I_q, b) \tag{3}$$

where N represents the number of pixels in image I, $\gamma \in (0, 1)$ is a parameter used for controlling the weight reduction according to the distance between q and p. I_q is the intensity value of pixel q, and the value of $Q(I_q, b)$ is zero except when I_q falls into bin b. If the image I is 1D, $H_p^E(b)$ can be computed efficiently using the following equation:

$$H_p^E(b) = H_p^{E,left}(b) + H_p^{E,right}(b) - Q(I_p, b) \tag{4}$$

where $H_p^{E,left}(b)$ denotes the LSH on the left of pixel p, and $H_p^{E,right}(b)$ is the LSH on the right of pixel p. The two LSHs are computed using the following equation:

$$H_p^{E,left}(b) = Q(I_p, b) + \gamma \cdot H_{p-1}^{E,left}(b) \tag{5}$$

$$H_p^{E,right}(b) = Q(I_p, b) + \gamma \cdot H_{p+1}^{E,right}(b) \tag{6}$$

Let H_p^O be the histogram for image region O_p centered at pixel p, and b_p is the bin that intensity value I_p falls in. According to the definition of histogram, we count the number of pixels in the image region O_p whose intensity values are within the interval $[b_p - e_p, b_p + e_p]$ as follow:

$$\mathbf{J}_p = \sum_{b=b_p-e_p}^{b_p+e_p} H_p^O(b) \tag{7}$$

where e_p denotes a parameter used for controlling the integration interval at pixel p. The integral value $\mathbf{J}_p$ represents a statistical feature of the image region O_p. Under the assumption of affine illumination transform, it is practically inaccurate to find an exact image region within which the affine illumination transform keeps invariant [58]. We hence adaptively consider the contribution of all the pixels in the image region O_p. Then, we replace the histogram H_p^O in (7) by the LSH H_p^E and (7) becomes:

$$\mathbf{J}_p = \sum_{b=1}^{B} exp\left(-\frac{(b-b_p)^2}{2max(\beta, e_p)^2} \cdot H_p^E(b)\right) \tag{8}$$

where $\beta = 0.1$ is a constant, $e_p = \beta|I_p - \bar{I}_p|$. $\bar{I}_p$ is the mean intensity value of all the pixels in image region O_p: $\bar{I}_p = \frac{1}{|o_p|}\sum_{q \in O_p} I_q$, and $|o_p|$ is the total pixel number in region O_p.

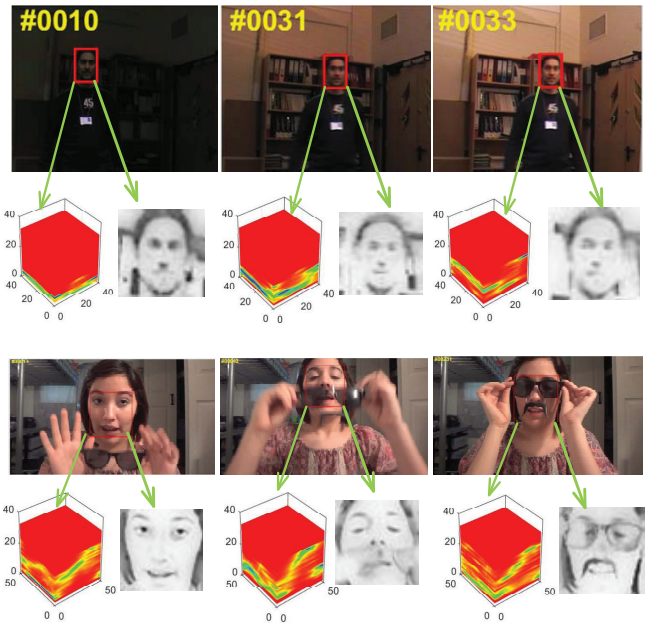

Fig. 2: LSH features of exmple frames.

In (8), $\mathbf{J}_p$ shows an illumination invariant feature computed on the basis of LSH which is different from the intensity value. $\mathbf{J}_p$ keeps invariant even under severe illumination changes (for more details of LSH, please refer to [43]). Figure 2 show examples of LSHs features for two image sequences: *Man* and *sunglasses*005. We use this LSH-based feature as foreground feature in our tracking method.

C. Objective Function of Our CF Model

In our work, we aim to learn a content aware CF (CACF) for face tracking. The objective function of the tracker is defined as follow:

$$\begin{aligned}\varepsilon(\mathbf{h}) &= \frac{1}{2}\sum_{j=1}^{T}\left\|\mathbf{y}(j) - \sum_{k=1}^{K}(\mathbf{h}\odot\mathbf{J})_k{}^{\top}\mathbf{P}\mathbf{x}_k[\triangle\tau_j]\right\|_2^2 \\ &+ \frac{\gamma}{2}\sum_{k=1}^{K}\|(\mathbf{h}\odot\mathbf{J})_k\|_2^2\end{aligned} \tag{9}$$

In this equation, $\mathbf{P}$ is a binary matrix used for cropping the mid D elements of a given signal $\mathbf{x}_k \in R^T$, satisfying $T \gg D$. The size of $\mathbf{P}$ is $D \times T$. The $\odot$ operator denotes the element-wise product. $\mathbf{h} = [h_1, h_2, ..., h_k]$ is the correlation filter $\mathbf{h} \in R^D$. $\mathbf{J} \in R^D$ is the LSH based feature. $\mathbf{y} \in R^T$ is the correlation output.

The training sample $\mathbf{x}$ are circularly shifted and then the cropping operator $\mathbf{P}$ is applied to crop the shifted training sample to obtain desired patches with size D from current frame. Those patches that correspond to the peak of $\mathbf{y}$ are positive examples showing the target of interest while the patches corresponding to the zero values of $\mathbf{y}$ are negative samples showing the background content. In order to further enhance the content of the

target of interest, the LSH based feature is incorporated into the objective function. When the appearance of the target is changed, the LSH based feature can benefit the filter to avoid possible model drift, which does greatly improve the ability of the filters in dealing with appearance variation induced by out of view, low resolution and in plane rotation.

D. Optimization of Our CF Model

In order to solve (9) efficiently, similar to the typical CF trackers, we can convert it into the frequency domain which is expressed as follow:

$$\begin{aligned} \varepsilon(\mathbf{h},\hat{\mathbf{g}}) &= \frac{1}{2}\left\|\hat{\mathbf{y}}-\hat{\mathbf{X}}\hat{\mathbf{g}}\right\|_2^2+\frac{\gamma}{2}\|\mathbf{h}\odot\mathbf{J}\|_2^2 \\ s.t. &\quad \hat{\mathbf{g}}=\sqrt{T}(\mathbf{F}\mathbf{P}^\top\otimes\mathbf{I}_K)*(\mathbf{h}\odot\mathbf{J}) \end{aligned} \tag{10}$$

where $\hat{\mathbf{g}}$ denotes an auxiliary variable to shorten the expression, and we define the attached matrix $\hat{\mathbf{X}} = [diag(\hat{\mathbf{x}}_1)^\top,...,diag(\hat{\mathbf{x}}_K)^\top]$ and its size is $T\times KT$. $\mathbf{h} = [\hat{h}_1^\top,...,\hat{h}_K^\top]$, $\hat{\mathbf{g}} = [\hat{\mathbf{g}}_1^\top,...,\hat{\mathbf{g}}_K^\top]$ and $\mathbf{I}_K$ is a $K\times K$ identity matrix. The symbol $\otimes$ denotes the Kronecker product and $\hat{}$ represents the discrete Fourier transform of a given signal, such that $\hat{\mathbf{h}}=\sqrt{T}\mathbf{F}\mathbf{h}$, where $\mathbf{F}$ represents an orthonormal matrix of complex basis vectors that is used for transforming any T dimensional vectorized signal into the Fourier domain. The size of $\mathbf{F}$ is $T\times T$. We use a symbol $\mho$ to represent $\mathbf{h}\odot\mathbf{J}$.

The ADMM method [44] is adopted to find the optimal solution of the CACF model. First, we can obtain the augmented Lagrangian form of (10) as follow:

$$\begin{aligned} Ł(\hat{\mathbf{g}},\mho,\hat{\ell}) &= \frac{1}{2}\left\|\hat{\mathbf{y}}-\hat{\mathbf{X}}\hat{\mathbf{g}}\right\|_2^2+\frac{\gamma}{2}\|\mho\|_2^2 \\ &+ \hat{\ell}^\top(\hat{\mathbf{g}}-\sqrt{T}(\mathbf{F}\mathbf{P}^\top\otimes\mathbf{I}_K)\mho) \\ &+ \frac{\omega}{2}\left\|\hat{\mathbf{g}}-\sqrt{T}(\mathbf{F}\mathbf{P}^\top\otimes\mathbf{I}_K)\mho\right\|_2^2 \end{aligned} \tag{11}$$

where $\hat{\ell}=[\hat{\ell}_1^\top,...,\hat{\ell}_K^\top]^\top$ denotes the Lagrangian vector defined in the Fourier domain with size $KT\times 1$, the symbol ω is a regularization constant. Then the ADMM method is adopted to solve this equation. It will convert the complex formulas to two subproblems alternatively in order to obtain a closed solution.

Subproblem $\mho$:

$$\begin{aligned} \mho &= \arg\min_{\mho}\left\{\frac{\gamma}{2}\|\mho\|_2^2+\hat{\ell}^\top(\hat{\mathbf{g}}-\sqrt{T}(\mathbf{F}\mathbf{P}^\top\otimes\mathbf{I}_K)\mho)\right. \\ &\left.+\frac{\omega}{2}\left\|\hat{\mathbf{g}}-\sqrt{T}(\mathbf{F}\mathbf{P}^\top\otimes\mathbf{I}_K)\mho\right\|_2^2\right\} \end{aligned} \tag{12}$$

Solve the partial derivative of $\mho$:

$$\frac{\partial Ł}{\partial\mho}=\gamma\mho-T\ell-\omega T\mathbf{g}+\mu T\mho=0 \tag{13}$$

So,

$$\mho=\left(\omega+\frac{\gamma}{T}\right)^{-1}(\omega\mathbf{g}+\ell) \tag{14}$$

where $\mathbf{g}=\frac{1}{\sqrt{T}}(\mathbf{P}\mathbf{F}^\top\otimes\mathbf{I}_K)\hat{\mathbf{g}}$, and $\ell=\frac{1}{\sqrt{T}}(\mathbf{P}\mathbf{F}^\top\otimes\mathbf{I}_K)\hat{\ell}$. Then $\mathbf{g}$ and ℓ can be divided into K independent IFFT calculations of $\mathbf{g}=\frac{1}{\sqrt{T}}\mathbf{P}\mathbf{F}^\top\hat{\mathbf{g}}$ and $\ell=\frac{1}{\sqrt{T}}\mathbf{P}\mathbf{F}^\top\hat{\ell}$. Furthermore, both $\mathbf{g}_k$ and ℓ are computed efficiently using IFFT on each $\hat{\mathbf{g}}$ and $\hat{\ell}$ ($\mathbf{g}_k=\frac{1}{\sqrt{T}}\mathbf{F}^\top\hat{\mathbf{g}_k}$, $\ell_k=\frac{1}{\sqrt{T}}\mathbf{F}^\top\hat{\ell}_k$). Thus, the computation complexity is $\mathcal{O}(KT\log T)$.

Subproblem g:

$$\begin{aligned} \hat{\mathbf{g}} &= \arg\min_{\hat{\mathbf{g}}}\left\{\frac{1}{2}\left\|\hat{\mathbf{y}}-\hat{\mathbf{X}}\hat{\mathbf{g}}\right\|_2^2\right. \\ &+ \hat{\ell}^\top(\hat{\mathbf{g}}-\sqrt{T}(\mathbf{F}\mathbf{P}^\top\otimes\mathbf{I}_K)\mho) \\ &\left.+\frac{\omega}{2}\left\|\hat{\mathbf{g}}-\sqrt{T}(\mathbf{F}\mathbf{P}^\top\otimes\mathbf{I}_K)\mho\right\|_2^2\right\} \end{aligned} \tag{15}$$

Due to its computation complexity, solving (15) to achieve real-time tracking is extremely difficult. Since the value of each pixel is independent, we consider to solve for $\hat{\mathbf{g}}$ at all the iterations of ADMM. We can represent $\hat{\mathbf{g}}$ as T separate objectives $\mathbf{g}(t)$: $\hat{\mathbf{g}}(t) = [conj(\hat{\mathbf{g}}_1(t)),...,conj(\hat{\mathbf{g}}_K(t))]^\top$. The operator $conj(.)$ refers to the complex conjugate operator of a complex vector. So (15) can be reformulated as :

$$\begin{aligned} \hat{\mathbf{g}}(t) &= \arg\min_{\hat{\mathbf{g}}(t)}\left\{\frac{1}{2}\left\|\hat{\mathbf{y}}(t)-\hat{\mathbf{x}}(t)^\top\hat{\mathbf{g}}(t)\right\|_2^2\right. \\ &+ \hat{\ell}(t)^\top(\hat{\mathbf{g}}(t)-\hat{\mho}(t)) \\ &\left.+\frac{\omega}{2}\left\|\hat{\mathbf{g}}(t)-\hat{\mho}(t)\right\|_2^2\right\} \end{aligned} \tag{16}$$

where $\hat{x}(t)=[\hat{x}_1(t),...,\hat{x}_K(t)]$, $\hat{\mho}_K(t)=\sqrt{T}\mathbf{F}\mathbf{P}^\top\mho$, $\hat{\mho}(t)=[\hat{\mho}_1(t),...,\hat{\mho}_K(t)]$. Similar to the solution of (12), the solution of $\hat{\mathbf{g}}(t)$ can be obtained by:

$$\begin{aligned} \hat{\mathbf{g}}(t) &= (\hat{\mathbf{x}}(t)\hat{\mathbf{x}}(t)^\top+T\omega\mathbf{I}_K)^{-1} \\ &\quad(\hat{\mathbf{y}}(t)\hat{\mathbf{x}}(t)-T\hat{\ell}(t)+T\omega\hat{\mho}(t)) \end{aligned} \tag{17}$$

Even if we acquire (15) for $\hat{\mathbf{g}}(t)$, the calculation is still a puzzle because we need real-time tracking. So we utilize the Sherman-Morrison formula to accelerate computation: $(\mathbf{A}+\mathbf{u}\mathbf{v}^\top)^{-1}=\mathbf{A}^{-1}-\frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^\top\mathbf{A}^{-1}}{1+\mathbf{v}^\top\mathbf{A}^{-1}\mathbf{u}}$, where in our model, $\mathbf{A}=T\omega\mathbf{I}_K$ and $\mathbf{u}=\mathbf{v}=\hat{\mathbf{x}}(t)$. Therefore, we rewrite (15) as:

$$\begin{aligned} \hat{\mathbf{g}}(t) &= \frac{1}{\omega T}\left(\hat{\mathbf{y}}(t)\hat{\mathbf{x}}(t)-T\hat{\ell}(t)+\omega T\hat{\mho}(t)\right) \\ &\quad-\frac{\hat{\mathbf{x}}(t)}{\omega Te}\left(\hat{\mathbf{y}}(t)\hat{s}_{\mathbf{x}}(t)-T\hat{s}_\ell(t)+\omega T\hat{s}_\mho(t)\right) \end{aligned} \tag{18}$$

where, $\hat{s}_{\mathbf{x}}(t)=\hat{\mathbf{x}}(t)^\top\hat{\mathbf{x}}$, $\hat{s}_\ell(t)=\hat{\mathbf{x}}(t)^\top\hat{\ell}$, $\hat{s}_\mho(t)=\hat{\mathbf{x}}(t)^\top\hat{\mho}$, and $e=\hat{s}_{\mathbf{x}}(t)+T\omega$.

Lagrangian Multiplier Update:

$$\hat{\ell}^{(i+1)} = \hat{\ell}^{(i)}+\omega(\hat{\mathbf{g}}^{(i+1)}-\hat{\mho}^{(i+1)}) \tag{19}$$

where $\hat{\ell}^{(i)}$ represents the Fourier transform of the Lagrangain in the previous state. $\hat{\mathbf{g}}^{(i+1)}$ and $\hat{\mho}^{(i+1)}$ are

the present solutions to the above subproblems at iteration $i+1$ within the iterative ADMM period. It is worth mentioned that ω is usually set to be $\omega^{(i+1)} = \min(\omega_{\max}, \beta\omega^{(i)})$.

Online Update: We adopt an online adaptation scheme which is similar to conventional CF trackers, such as [35, 59], in order to further boost the performance of our tracker. At frame f, the model adaptation is formulated as:

$$\hat{\mathbf{x}}_{model}^{(f)} = (1-\varphi)\hat{\mathbf{x}}_{model}^{(f-1)} + \varphi\hat{\mathbf{x}}^{(f)} \tag{20}$$

The parameter φ is adaptation rate. In (18), $\hat{\mathbf{x}}_{model}^{(f)}$ will be used to compute the corresponding terms.

Object Localization The final step is to determine the position of the target face which is detected by applying the updated filter of frame $f-1$: $\hat{\mathbf{g}}^{(f-1)}$. The spatial location of the target face can be computed in the Fourier domain using the following equation:

$$\hat{\mathbf{r}} = \sum_{k=1}^{K} \hat{\mathbf{x}}_{\mathbf{k}} \odot \hat{\mathbf{g}}_{\mathbf{k}} \tag{21}$$

where $\mathbf{r}$ is the response map and $\hat{\mathbf{r}}$ denotes its Fourier transform. After the response map is obtained, the promising face location is obtained based on the maximum response.

Fig. 3 shows an illustration of the proposed content aware CF.

IV. Experiment

In this section, we present the experimental result of the CACF tracker. We first describe the face tracking dataset collected for performance evaluation. Then, we demonstrate the effectiveness of our tracker compared with several advanced tracking methods.

Our method is implemented in MATLAB and tested on a tower workstation with CPU Intel Core i7-9700 3.0GHz and 48GB RAM. In our experiment, we adopt 31-channel HOG features with each feature cell size 4×4, which is commonly used in [36, 38, 59]. The regularization weight γ is set to be 0.01. The learning rate φ is 0.013. The ADMM iteration number is 2 and the parameter ω is set as 1, while parameter β for updating ω is 4 and $\omega_{\max}$ is 10000. For LSH feature, as in [43], the number of bins is B =32 and the parameter $\alpha = 0.15$, while the parameter in (8) is 0.1.

We compare the proposed CACF tracker with several state-of-the-art tracking methods including CF trackers and deep trackers: SiamRPN++ [60], DASiamRPN [61], SiamMask [62], AutoTrack [63], SKSCF [64], BACF [40], ECO_HC [65], Staple [59], SRDCF [38], fDSST [45], DSST [66], SAMF_CA [39].

A. Evaluation Metrics

Following the standard paradigm in object tracking, we use the success and precision plots mentioned in [9, 67] to evaluate all the trackers. All of them are ranked based on the area under curve scores (AUC) of their success plots. The precision plots are generated based on the trackers' center location errors (CLE), which is defined as the average Euclidean distance between the estimated face location center and ground-truth center. In our experiments, we use 20 pixels as the CLE threshold for ranking trackers.

B. Collected Face Dateset

In order to do performance evaluation of the trackers, we collect 97 face video sequences and manually annotate the dataset according to public standard. The challenging dataset contains different kinds of challenging attributes in general object tracking. To the best of our knowledge, our dataset the largest face tracking dataset that contains both indoor and outdoor faces in the literature. We name the dataset as **FaceSet**. The video sequences in FaceSet are collected from four visual tracking datasets: OTB100 [67], ClemsonHeadSeqs [68] and NUS-PRO [69] and BUAA-PRO [70]. Table 1 shows the detailed challenges and resolution of each face sequence.

C. Quantitative Results

The success and precision plots of our tracker against state-of-the-art trackers are shown in figure 4. It is clear that our proposed CACF tracker gains the best performance in terms of precision score (0.898), while the runner-up is BACF (0.897). The deep trackers do not obtain satisfactory results over the FaceSet. The best deep tracker is SiamRPN++_mobile (0.763) which is 23.5% lower than the proposed CACF tracker. AutoTrack, which is very impressive in UAV object tracking, gets a precision score 0.842 that is 6.24% lower than CACF. In light of the success plots, the AUC score of CACF tracker (0.739) is the runner-up which is slightly less than the SRDCF (0.740). When compared to the baseline BACF tacker which is the second runner-up, our CACF obtains 0.004 improvement in terms of the AUC score. It is notable that all the deep trackers do not perform well enough on the FaceSet compared to the CF based trackers.

We also give the attribute-based results in figure 5. From figure 5, it is clear that the winners and runner-ups of different attributes are varied. In terms of precision plots, CACF ranks first in BC and IPR subsets, second in DEF, IV and SV subsets, and third in OCC subset. In OPR subset, the Siamese network based trackers occupy the top five positions, which demonstrates that they can cope with out-of-plane rotation better than CF based trackers. As for the success plots, CACF wins the championship in BC, DEF, IPR and SV subsets. In MB subset, CACF AUC score is 0.630 which is 9.76% higher than BACF (0.574). When it comes to the OCC subset, CACF (0.695) is the second runner-up followed by BACF (0.688), while SRDCF and ECO_HC ranks top two. For OPR subset, the results are similar to that of precision plots.

From the above analysis, it is clear that the overall performance of our CACF tracker is much stabler than its counterparts and is very competitive under different challenging attributes. When compared with the base-

Table 1: FaceSet sequences and challenges

Name	Resolution	Challenges	Name	Resolution	Challenges
Biker	640 × 360	OPR,SV,OCC,MB,FM,OV,LR	interview003	1280 × 720	IV,DEF,IPR
BlurFace	640 × 480	MB,FM,IPR	interview004	1280 × 720	DEF,IPR,BC
Boy	640 × 480	OPR,SV,MB,FM,IPR	interview005	1280 × 720	MB,IPR,BC
David	320 × 240	IV,OPR,SV,OCC,DEF,MB,IPR	interview006	1280 × 720	IV,SV,DEF
David2	320 × 240	OPR,IPR	interview007	1280 × 720	SV
DragonBaby	640 × 360	OPR,SV,OCC,MB,FM,IPR,OV	interview008	1280 × 720	SV,IPR
Dudek	720 × 480	OPR,SV,OCC,DEF,FM,IPR,OV,BC	interview009	1280 × 720	SV,BC
FaceOcc1	352 × 288	OCC	interview010	1280 × 720	SV,IPR
FaceOcc2	320 × 240	IV,OPR,OCC	interview011	1280 × 720	SV,IPR,BC
FleetFace	720 × 480	OPR,SV,DEF,MB,FM,IPR	interview012	1280 × 720	IV,SV,BC
Freeman1	360 × 240	OPR,SV,IPR	interview013	1280 × 720	SV
Freeman3	360 × 240	OPR,SV,IPR	interview014	1280 × 720	IV,SV,OCC,IPR
Freeman4	360 × 240	OPR,SV,OCC,IPR	interview015	1280 × 720	SV,BC
Girl	129 × 96	OPR,SV,OCC,IPR	interview016	1280 × 720	SV,IPR,BC
Jumping	352 × 288	MB,FM	interview017	1280 × 720	IV,SV,IPR,BC
KiteSurf	480 × 270	IV,OPR,OCC,IPR	interview018	1280 × 720	DEF,IPR
Man	241 × 193	IV	interview019	1280 × 720	SV
Mhyang	320 × 240	IV,OPR,DEF,BC	interview020	1280 × 720	IV,SV,OCC,DEF
Shaking	624 × 352	IV,OPR,SV,IPR,BC	politician001	1280 × 720	DEF,IPR
Soccer	640 × 360	IV,OPR,SV,OCC,MB,FM,IPR,BC	politician002	1280 × 720	IV,IPR
Surfer	480 × 360	OPR,SV,FM,IPR,LR	politician003	1280 × 720	IV,IPR
Trellis	320 × 240	IV,OPR,SV,IPR,BC	politician004	1280 × 720	IV,IPR
seqBB	128 × 96	IV,OPR,SV,IPR,BC	politician005	1280 × 720	IV,IPR
seqCubicle	128 × 96	IV,SV,OCC,IPR	politician006	1280 × 720	IV,OCC,IPR
seqDhb	128 × 96	IV,IPR	politician007	1280 × 720	IV,IPR
seqDjb	128 × 96	IV,SV,IPR,BC	politician008	1280 × 720	IPR
seqDk	128 × 96	IV,IPR	politician009	1280 × 720	IPR
seqDp	128 × 96	IV,SV,OCC,IPR,BC	politician010	1280 × 720	IPR
seqDt	128 × 96	IV,SV,DEF,IPR	sunglasses001	1280 × 720	IV,SV,OCC
seqFast	128 × 96	FM	sunglasses002	1280 × 720	IV,OCC,IPR
seqJd	128 × 96	IV,SV,OCC,IPR,BC	sunglasses003	1280 × 720	IV,OCC,IPR
seqMg	128 × 96	IV,SV,OCC,IPR,BC	sunglasses004	1280 × 720	IV,OCC
seqMs	128 × 96	OCC	sunglasses005	1280 × 720	IV,SV,OCC
seqSb	128 × 96	IV,SV,OCC,DEF,IPR	sunglasses006	1280 × 720	SV,OCC
seqSim	128 × 96	IV,SV,OCC,IPR	sunglasses007	1280 × 720	SV,OCC
seqVillains1	128 × 96	IV,SV,OCC,IPR,BC	sunglasses008	1280 × 720	OCC,IPR
seqVillains2	128 × 96	IV,SV,OCC,IPR,OV,BC	sunglasses009	1280 × 720	OCC
hat001	1280 × 720	IV,OCC,IPR	sunglasses010	1280 × 720	OCC,IPR
hat002	1280 × 720	IV,OCC,IPR	mask001	1280 × 720	IV,OCC,DEF,IPR
hat003	1280 × 720	IV,OCC,MB,IPR	mask002	1280 × 720	OCC,IPR
hat004	1280 × 720	IV,OCC	mask003	1280 × 720	OCC,IPR
hat005	1280 × 720	OCC,MB,IPR	mask004	1280 × 720	OCC,IPR
hat006	1280 × 720	SV,OCC,MB,IPR	mask005	1280 × 720	OCC,DEF,IPR
hat007	1280 × 720	IV,SV,OCC	mask006	1280 × 720	OCC
hat008	1280 × 720	IV,OCC	mask007	1280 × 720	OCC,IPR
hat009	1280 × 720	IV,SV,OCC,DEF	mask008	1280 × 720	SV,OCC,IPR
hat010	1280 × 720	IV,SV,OCC	mask009	1280 × 720	IV,OCC,IPR
interview001	1280 × 720	DEF,IPR	mask010	1280 × 720	IV,OCC,IPR
interview002	1280 × 720	IV,SV	–	–	–

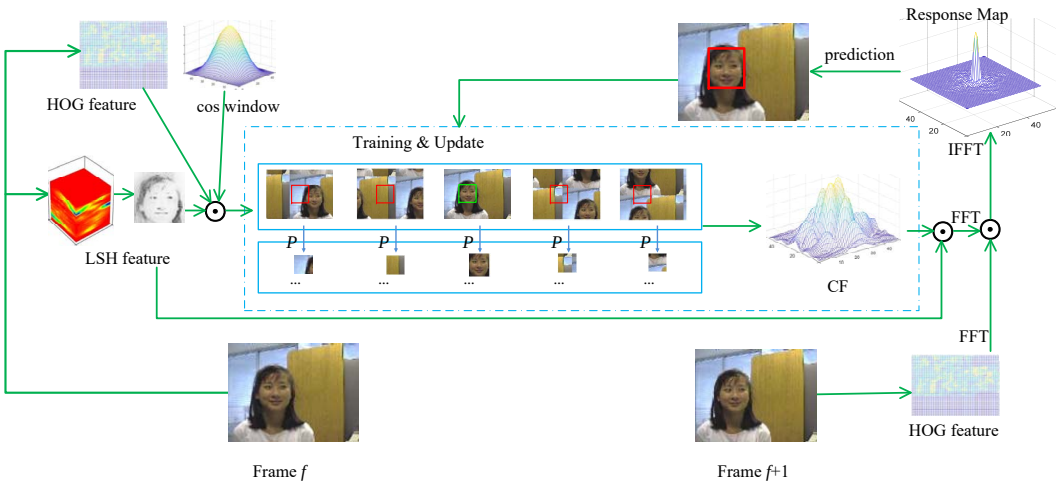

Fig. 3: Illustration of the proposed content aware correlation filter.

line CF trackers, especially the BACF tracker, our CACF could boost the tracking performance in almost all the challenging situations. This is mainly attributed to the incorporation of both foreground and background information within the CF framework.

D. Qualitative Results

We select 6 typical face sequences from FaceSet to show the qualitative results, that are David, hat001, hat003, hat006, mask006 and sunglasses005. Sample tracking results of five CF trackers are shown in figure 6 (CACF, BACF, ECO_HC, SiamMask and SiamRPN++_r50).

In the first row (David), it shows that our tracker can capture the face successfully under severe illumination change and in plane rotation. The faces in the second (hat001), third (hat003) and fourth (hat005) rows undergo frequent rotation, deformation and occlusion, but our CACF can still track the faces accurately under these challenging factors. In the fifth and sixth rows, the faces are interrupted by heavy makeups and decorations which makes it difficult to capture them accurately. The fifth row corresponds to the *Mask*006 sequence while the sixth row *Sunglasses*005. The sample frames in the two rows show that almost all the five trackers can capture the faces, thus it is difficult to directly show the improved performance of our CACF tracker against the other methods. We compute the center location errors (CLE) of the trackers on these two sequences. The CLEs on these sample frames are listed in Table 2. It is clear that our CACF tracker is obviously better than the other four methods. BACF and ECO_HC are better than SiamRPN++_r50 and SiamMask.

V. Discussion

As we analyzed in the experiment section, the proposed CACF could boost the performance of face tracking under different challenging situations. The main advantage of CACF is the real negative training samples drawn from the background patches and the incorporation of the locality sensitive histogram based foreground features. For CF based trackers, traditional negative samples are mainly obtained from circular shifted foreground patches, which leads to annoying boundary effects. In order to suppress such effect, real negative samples must be used for training the CF. On the other hand, leveraging background information could boost the tracking performance and it is common in constructing trackers. As shown in figure 3, the background patches are sampled as the negative samples in the training block. In order to further increase the discriminative ability of the tracker, we adopt the LSH based feature as the foreground feature and incorporate it into the CF framework. By using the above strategies, the learned CF could highlight the target face area in the response map (see figure 3).

Despite the improved face tracking performance of CACF, the main limitation of the CACF lies in that the weight γ in the spatial regularization term of our objective function keep invariant during tracking, which cannot well adapt to the variation of target face. Thus, it is necessary to change the weight of spatial regularization according to different face appearances. Another limitation is the ADMM method for solving the objective function. Since the ADMM methods requires ergodic averaging of variables [44], it destroys the sparsity of the solution, leading that the convergence rate is not optimal[71]. This problem will ignite the exploration of improved ADMM method for solving the objective function of various CF trackers.

VI. Conclusion

In this paper, we have proposed a content aware correlation filter for face tracking. In order to solve the boundary effect in CF trackers, we exploit the background information for learning a CF in which the background patches are used as negative samples while the target patches as positive samples. A locality sensitive histogram based illumination invariant feature is used as foreground feature to discriminate the target face from complex background. The feature is incorporated into

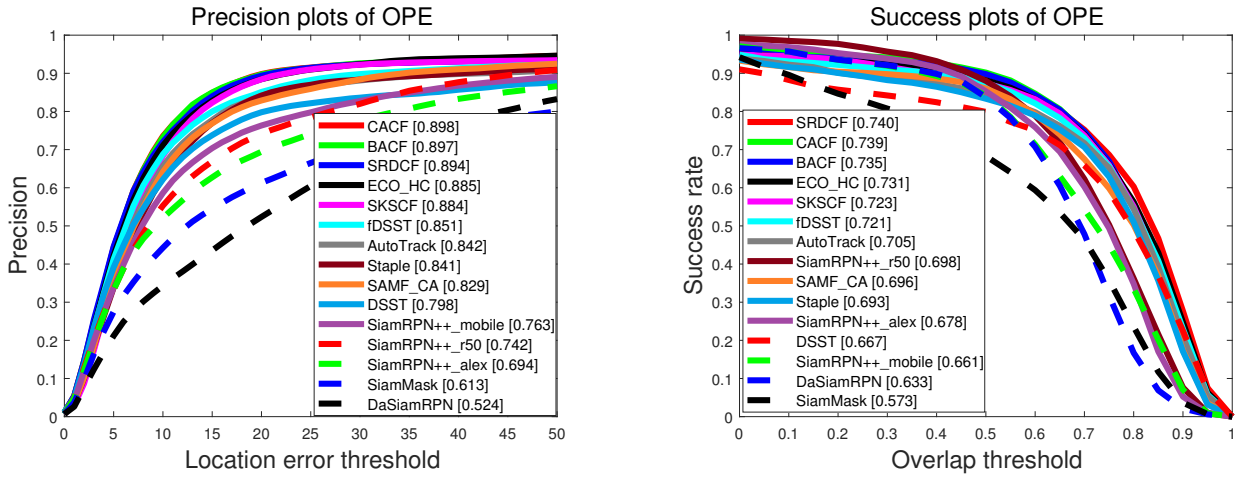

Fig. 4: Precision plots and success plots on FaceSet compared against state-of-the-art tracking methods.

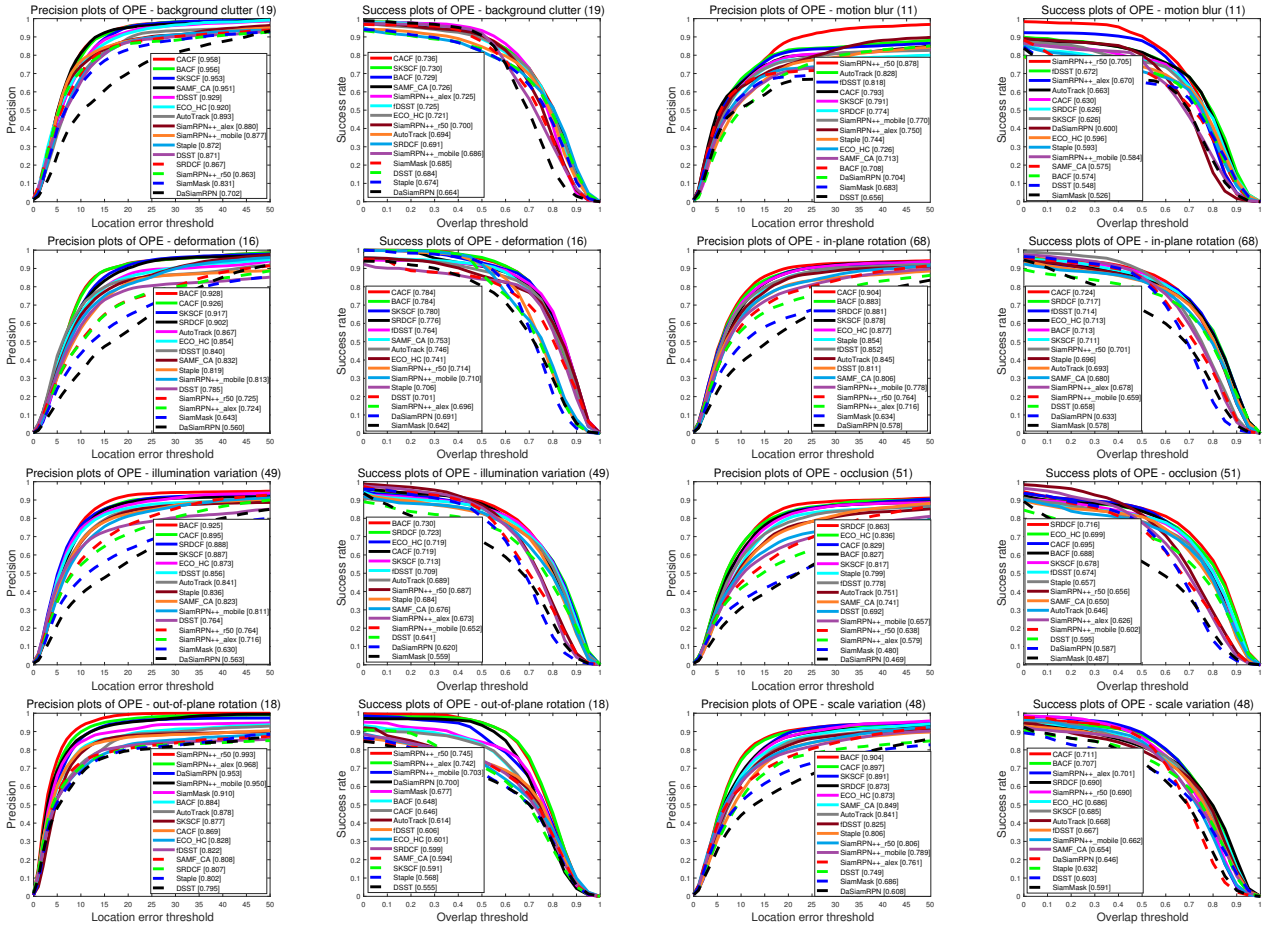

Fig. 5: Attribute-based comparison. We list the precision and success plots of eight attributes. The other three attributes, FM, LR and OV are not shown because the number of sequences containing the three attributes are 9, 2 and 4, respectively.

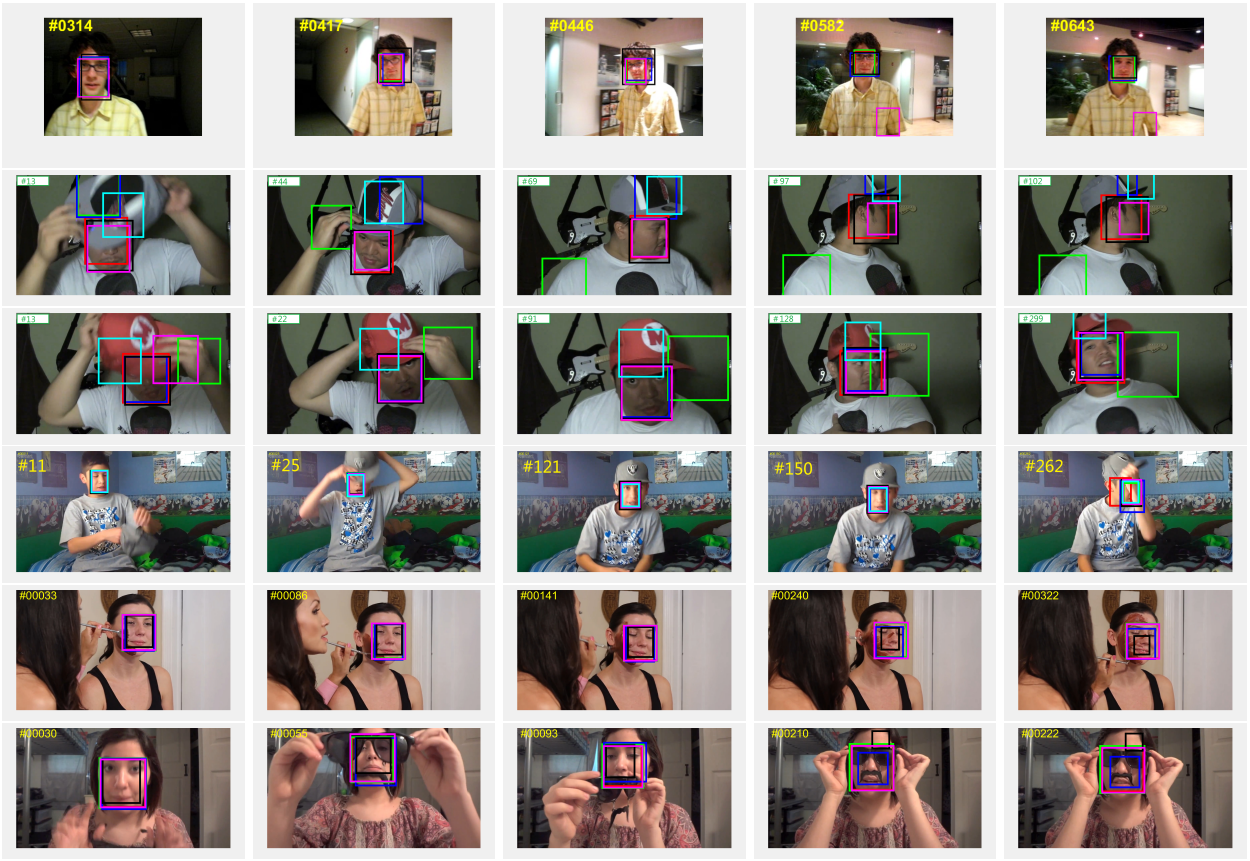

Fig. 6: Sample frames of tracking results over several face sequences from FaceSet. Red: CACF, Green: BACF, Blue: SiamRPN++_r50, Black: SiamMask, Pink: ECO_HC

Table 2: CLEs of the trackers on the sample frames of Fig. 6

Seq. # FrameNo.	CACF	BACF	SiamRPN++_r50	SiamMask	ECO_HC
Mask006#33	0.5	0.5	4.533	5.453	1.803
Mask006#86	0.707	1.118	2.625	12.834	2.828
Mask006#141	2.5	2.549	10.052	6.136	2.692
Mask006#240	5.148	5.701	18.037	22.255	5.523
Mask006#322	2.0	2.236	19.410	15.389	3.640
Sunglasses#030	6.042	6.50	13.481	12.362	7.211
Sunglasses#055	1.803	2.50	46.999	31.098	12.806
Sunglasses#093	7.433	8.515	37.667	33.056	8.016
Sunglasses#210	8.322	6.708	32.669	177.741	1.803
Sunglasses#222	2.236	2.550	18.564	176.217	2.915

the objective function and the ADMM method is used to solve the objective function. We also build a face tracking dataset (FaceSet) which contains 97 sequences and covers 11 challenging attributes in face tracking. The resulting content aware correlation filter shows promising performance improvement compared to other CF based trackers in the FaceSet. In our future work, we will further explore the potential of the content aware CF framework to incorporate deep features for boosted performance and expand the FaceSet with more sequences. We will also try to improve the CACF tracker by introducing adaptive spatial regularization weight and accelerated ADMM method to further boost the face tracking performance.

References

[1] S. Yeung, F. Rinaldo, J. Jopling, B. Liu, R. Mehra, N. L. Downing, M. Guo, G. Bianconi, A. Alahi, J. Lee, B. Campbell, K. Deru, W. Beninati, F.-f. Li, and A. Milstein, "A computer vision system for deep learning-based detection of patient mobilization activities in the icu," *npj Digit. Med.*, vol. 2, pp. 1–5, 2019.

[2] K. Adapa, S. Jain, R. Kanwar, T. Zaman, T. Taneja, J. Walker, and L. Mazur, "Augmented reality in patient education and health literacy: a scoping review protocol," *BMJ Open*, vol. 310, no. e038416, pp. 1–9, 2020.

[3] Y. J. Lee and Y. J. Lee, "Face tracking for augmented reality game interface and brand placement," in *UCMA 2011: Ubiquitous Computing and Multimedia Applications*, ser. Communications in Computer and Information Science, T. Kim, H. Adeli, R. Robles, and M. Balitanas, Eds. Springer, 2011, vol. 151, pp. 72–78.

[4] P. Gupta, B. Bhowmick, and A. Pal, "Mombat: heart rate monitoring from face video using pulse modeling and bayesian tracking," *Comput. Biol. Med.*, vol. 121, p. 103813, 2020.

[5] V. Srisamosorn, N. Kuwahara, A. Yamashita, T. Ogata, and J. Ota, "Design of face tracking system using environmental cameras and flying robot for evaluation of health care," in *DHM 2016: Digital Human Modeling: Applications in Health, Safety, Ergonomics and Risk Management*, ser. Lecture Notes in Computer Science, V. Duffy, Ed. Springer, June 2016, vol. 9745, pp. 264–273.

[6] P. Huber, P. Kopp, W. Christmas, M. Rtsch, and J. Kittler, "Real-time 3d face fitting and texture fusion on in-the-wild videos," *IEEE Signal Proc. Let.*, vol. 24, no. 4, pp. 437–441, 2017.

[7] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Niebner, "Facevr: real-time gaze-aware facial reenactment in virtual reality," *ACM T. Graphic.*, vol. 37, no. 2, pp. 1–15, June 2018.

[8] A. K. Roy-Chowdhury and Y. Xu, *Face tracking*. Boston, MA: Springer US, 2015, pp. 532–538.

[9] Y. Wu, J. Lim, and M.-H. Yang, "Online Object Tracking: A Benchmark," in *Proc. CVPR*. IEEE, Jun. 2013, pp. 2411–2418.

[10] G. Chrysos, E. Antonakos, P. Snape, A. Asthana, and S. Zafeiriou, "A comprehensive performance evaluation of deformable face tracking "in-the-wild"," *Int. J. Comput. Vision*, vol. 126, pp. 198–232, 2018.

[11] J. Shen, S. Zafeiriou, G. G. Chrysos, J. Kossaifi, G. Tziiropoulos, and M. Pantic, "The first facial landmark tracking in-the-wild challenge: benchmark and results," in *Proc. ICCVW*. IEEE, Dec. 2015, pp. 1003—1011.

[12] U. Prabhu, K. Seshadri, and M. Savvides, "Automatic facial landmark tracking in video sequences using kalman filter assisted active shape models," in *Trends and Topics in Computer Vision*, K. N. Kutulakos, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 86–99.

[13] H. Kim, H. Kim, and E. Hwang, "Real-time shape tracking of facial landmarks," *Multimed. Tools Appl.*, vol. 79, pp. 15 945–15 963, 2020.

[14] V. Contreras-Gonzalez, V. H. Diaz-Ramirez, and R. Juarez-Salazar, "Facial landmark detection and tracking with dynamically adaptive matched filters," *J. Electron. Imaging*, vol. 29, no. 3, pp. 033 004.1–18, 2020.

[15] H. K. Almohair, "An icsc model for detecting human skin in jpeg images," *WSEAS Trans. Signal Process.*, vol. 16, pp. 75–80, 2020.

[16] A. Bulbul, Z. Cipiloglu, and T. Capin, "A color-based face tracking algorithm for enhancing interaction with mobile devices," *Visual Comput.*, vol. 26, pp. 311–323, 2010.

[17] M. Goyani, G. Shikkenawis, and B. Joshi, "Geometry and skin color based hybrid approach for face tracking in colour environment," in *CCSIT 2011: Advances in Computer Science and Information Technology*, ser. Communications in Computer and Information Science, N. Meghanathan, B. Kaushik, and D. Nagamalai, Eds. Springer, 2011, vol. 131, pp. 339–347.

[18] J.-H. Kim, B.-D. Kang, J.-S. Eom, C.-S. Kim, S.-H. Ahn, B.-J. Shin, and S.-K. Kim, "Real-time face tracking system using adaptive face detector and kalman filter," in *HCI 2007: Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments*, ser. Lecture Notes in Computer Science, J. Jacko, Ed. Springer, 2007, vol. 4552, pp. 669–678.

[19] Q. N. Vo and G. Lee, "A feature-based adaptive model for realtime face tracking on smart phones," in *SCIA 2013: Image Analysis*, ser. Lecture Notes in Computer Science, J.-K. Kamarainen and M. Koskela, Eds. Springer, 2013, vol. 7944, pp. 630–639.

[20] V. Varadarajan, S. Lokesh, A. Ramesh, A. Vanitha, and V. Vaidehi, "Face tracking using modified forward-backward mean-shift algorithm," in *DaSAA 2017: Data Science Analytics and Applications*, ser. Communications in Computer and In-

formation Science, R. Shriram and M. Sharma, Eds. Springer, 2017, vol. 804, pp. 46–59.

[21] Y.-S. Huang and C.-I. Chang, "Multi-face tracking with occlusion recovery," in *Proc. ICGEC*. Springer, Aug. 2015, pp. 247—257.

[22] K.-Y. Liu, Y.-H. Li, S. Li, L. Tang, and L. Wang, "A new parallel particle filter face tracking method based on heterogeneous system," *J. Real-Time Image Proc.*, vol. 7, pp. 153–163, 2012.

[23] Y.-H. Lee, M.-H. Jeong, J.-J. Lee, and B.-J. You, "Robust face tracking using bilateral filtering," in *ICIC 2008: Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues*, ser. Lecture Notes in Computer Science, D.-S. Huang, Ed. Springer, 2008, vol. 5226, pp. 1181–1189.

[24] K. Zhang, E. Barati, E. Rashedi, and X. Chen, "Long-term face tracking in the wild using deep learning," in *Proc. KDD Workshop on Large-scale Deep Learning for Data Mining*, Aug. 2016, pp. 1–13.

[25] Y. Qi, S. Zhang, F. Jiang, H. Zhou, and D. Tao, "Siamese local and global networks for robust face tracking," *IEEE Trans. Image Process.*, vol. 29, pp. 9152 – 9164, 2020.

[26] X. Li and J. Lang, "Simple real-time multi-face tracking based on convolutional neural networks," in *Proc. ICCRV*. IEEE, May. 2018, pp. 337–344.

[27] Z. Lian, S. Shao, and C. Huang, "A real time face tracking system based on multiple information fusion," *Multimed. Tools Appl.*, vol. 79, pp. 16 751–16 769, 2020.

[28] M. Gu, J. Lu, and J. Zhou, "Dual-agent deep reinforcement learning for deformable face tracking," in *Proc. ECCV*. Springer, Sep. 2018, pp. 783–799.

[29] A. Sleit, R. Abu-Hurra, and W. Almobaideen, "Lower-quarter-based face verification using correlation filter," *Imaging Sci. J.*, vol. 59, no. 1, pp. 41–48, 2011.

[30] X. Zhu, S. Liao, Z. Lei, R. Liu, and S. Z. Li, "Feature correlation filter for face recognition," in *Proc. ICB*. Springer, Sep. 2005, pp. 77–86.

[31] M. Taheri, "Robust face recognition via non-linear correlation filter bank," *IET Image Process.*, vol. 12, no. 3, pp. 408–415, 2017.

[32] V. D. My and A. Zell, "Real time face tracking and pose estimation using an adaptive correlation filter for human-robot interaction," in *Proc. ECMR*. IEEE, Sep. 2013, pp. 119–124.

[33] L. N. Gaxiola, V. H. Diaz-Ramirez, J. J. Tapia, A. Diaz-Ramirez, and V. Kober, "Robust face tracking with locally-adaptive correlation filtering," in *CIARP 2014: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, ser. Lecture Notes in Computer Science, E. Bayro-Corrochano and E. Hancock, Eds. Springer, Nov. 2014, vol. 8827, pp. 925–932.

[34] J. Su, L. Gao, W. Li, Y. Xia, N. Cao, and R. Wang, "Fast face tracking-by-detection algorithm for secure monitoring," *Appl. Sci. Basel*, vol. 9, pp. 1–17, 2019.

[35] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. CVPR*. IEEE, Jun. 2010, pp. 2544–2550.

[36] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, 2014.

[37] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *Proc. ECCV*. Springer, Sep. 2014, pp. 254–265.

[38] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proc. ICCV*. IEEE, Dec. 2015, pp. 4310–4318.

[39] M. Mueller, N. Smith, and B. Ghanem, "Context-aware correlation filter tracking," in *Proc. CVPR*. IEEE, Jun. 2017, pp. 1396–1404.

[40] H. Kiani Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in *Proc. ICCV*. IEEE, Oct. 2017, pp. 1135–1143.

[41] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg, "Convolutional features for correlation filter based visual tracking," in *Proc. ICCVW*. IEEE, Dec. 2015, pp. 58–66.

[42] Y. Lin, S. Cheng, J. Shen, and M. Pantic, "Mobiface: a novel dataset for mobile face tracking in the wild," in *Proc. FG*, May. 2019, pp. 1–8.

[43] S. He, Q. Yang, R. W. Lau, J. Wang, and M.-H. Yang, "Visual tracking via locality sensitive histograms," in *Proc. CVPR*. IEEE, Jun. 2013, pp. 2427–2434.

[44] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[45] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Discriminative scale space tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 8, pp. 1561–1575, 2017.

[46] J. Liao, Q. Wang, L. Cao, X. Jiahao, and Z. Yiting, "Mtcnn-kcf-deepsort: driver face detection and tracking algorithm based on cascaded kernel correlation filtering and deep sort," in *WCX SAE World Congress Experience*, ser. SAE Technical Paper. SAE, Apr. 2020, pp. 2020–01–1038.

[47] M. Soldic, D. Marcetic, and S. Ribaric, "A robust online multi-face tracking system," in *2018 International Symposium ELMAR*, Sep. 2018, pp. 159–163.

[48] S. M. Rathnam and G. Siva Koteswara Rao, "A novel deep learning architecture for image hiding," *WSEAS Trans. Signal Process.*, vol. 16, pp. 206–210, 2020.

[49] L. Males, D. Marcetic, and S. Ribaric, "A multi-agent dynamic system for robust multi-face track-

ing," *Expert Syst. Appl.*, vol. 126, pp. 246 – 264, 2019.

[50] W. W. Zou, P. C. Yuen, and R. Chellappa, "Low-resolution face tracker robust to illumination variations," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1726–1739, 2013.

[51] T. Chakravorty and E. Bilodeau, Guillaume-Alexandre ad Granger, "Robust face tracking using multiple appearance models and graph relational learning," *Mach. Vision Appl.*, vol. 31, no. 23, pp. 1–17, 2020.

[52] X. Jiang, H. Yu, Y. Lu, and H. Liu, "A fusion method for robust face tracking," *Multimed. Tools Appl.*, vol. 75, pp. 11 801–11 813, 2016.

[53] T. Li, P. Zhou, and H. Liu, "Multiple features fusion based video face tracking," *Multimed. Tools Appl.*, vol. 78, pp. 21 963–21 980, 2019.

[54] B. Wu, B.-G. Hu, and Q. Ji, "A coupled hidden markov random field model for simultaneous face clustering and tracking in videos," *Pattern Recogn.*, vol. 64, pp. 361 – 373, 2017.

[55] N. Le, A. Heili, D. Wu, and J.-M. Odobez, "Temporally subsampled detection for accurate and efficient face tracking and diarization," in *Proc. ICPR*. IEEE, Dec. 2016, pp. 1792–1797.

[56] D. Aspandi, O. Martinez, F. Sukno, and X. Binefa, "Fully end-to-end composite recurrent convolution network for deformable facial tracking in the wild," in *Proc. FG*. IEEE, Sep. 2019, pp. 1–8.

[57] D. Gordon, A. Farhadi, and D. Fox, "Re^3: real-time recurrent regression networks for visual tracking of generic objects," *IEEE Robot Autom. Let.*, vol. 3, no. 2, pp. 788–795, 2018.

[58] S. Chan, X. Zhou, J. Li, and S. Chen, "Adaptive compressive tracking based on locality sensitive histograms," *Pattern Recogn.*, vol. 72, pp. 517–531, 2017.

[59] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. Torr, "Staple: Complementary learners for real-time tracking," in *Proc. CVPR*. IEEE, Jun. 2016, pp. 1401–1409.

[60] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "Siamrpn++: Evolution of siamese visual tracking with very deep networks," in *Proc. CVPR*, 2019, pp. 4282–4291.

[61] Z. Zhu, Q. Wang, B. Li, W. Wu, J. Yan, and W. Hu, "Distractor-aware siamese networks for visual object tracking," in *Proc. ECCV*, 2018, pp. 101–117.

[62] Q. Wang, L. Zhang, L. Bertinetto, W. Hu, and P. H. Torr, "Fast online object tracking and segmentation: a unifying approach," in *Proc. CVPR*, Jun. 2019, pp. 1328–1338.

[63] Y. Li, C. Fu, F. Ding, Z. Huang, and G. Lu, "Autotrack: Towards high-performance visual tracking for uav with automatic spatio-temporal regularization," in *Proc. CVPR*, 2020, pp. 11 923–11 932.

[64] W. Zuo, X. Wu, L. Lin, L. Zhang, and M.-H. Yang, "Learning support correlation filters for visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 5, pp. 1158–1172, 2019.

[65] M. Danelljan, G. Bhat, F. Shahbaz Khan, and M. Felsberg, "Eco: Efficient convolution operators for tracking," in *Proc. CVPR*, Jun. 2017, pp. 6638–6646.

[66] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proc. BMVC*, 2014, pp. 1–11.

[67] Y. Wu, J. Lim, and M.-H. Yang, "Object Tracking Benchmark," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1834–1848, 2015.

[68] S. Birchfield, "Elliptical head tracking using intensity gradients and color histograms," in *Proc. CVPR*. IEEE, Jun. 1998, pp. 232–237.

[69] A. Li, M. Lin, Y. Wu, M. Yang, and S. Yan, "NUS-PRO: a new visual tracking challenge," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 335–349, 2016.

[70] A. Li, Z. Chen, and Y. Wang, "BUAA-PRO: a tracking dataset with pixel-level annotation," in *Proc. BMVC*, 2018.

[71] Y. Ouyang, Y. Chen, G. Lan, and E. P. Jr., "An accelerated linearized alternating direction method of multipliers," *SIAM J. Imaging Sci.*, vol. 8, no. 1, pp. 644–681, 2015.

Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

Houji Li and Fasheng Wang conceived this study.
Shuangshuang Yin was responsible for the collection of the dataset and exectuted the experiments.
Houjie Li and Fasheng Wang wrote the original manuscript.
Fuming Sun and Houjie Li revised the original manuscript.