Multimed Tools Appl (2016) 75:11801–11813
DOI 10.1007/s11042-015-2659-5

A fusion method for robust face tracking

Xiaodong Jiang [1] · **Hui Yu** [2] · **Yang Lu** [3] · **Honghai Liu** [2]

Received: 19 January 2015 / Revised: 1 April 2015 / Accepted: 23 April 2015 /
Published online: 3 May 2015
© Springer Science+Business Media New York 2015

Abstract Face tracking often encounters drifting problems, especially when a significant face appearance variation occurs. Many trackers suffer from the difficulty of facial feature extraction during a wide range of face turning, occlusion, and even invisibleness. In this paper, we propose a novel and efficient fusion strategy for robust face tracking. A Supervised Descent Method (SDM) and a Compressive Tracking method (CT) are employed at the same time. SDM is used to correct drifting errors of CT continuously during frontal face tracking. However, when the face orientation changes to the angle orthogonal to the view line, it results in tracking failure for the SDM method. CT is then adopted to keep the face region being tracked until SDM detects and tracks the face again. In the experiments, we test the proposed method for real-time tracking using several challenging sequences from recent literatures. The fusion strategy has achieved encouraging performance in terms of both efficiency and reliability.

Keywords Fusion algorithm · Human face tracking · Compressive tracking · Supervised descend method

1 Introduction

Face tracking is a primary step in computer vision due to its wide applications in robotic control, visual surveillance, video retrieval, human computer interaction and facial animation [4, 8, 19]. Although numerous approaches have been presented over the years, it is still a

✉ Xiaodong Jiang
xiaodong12jiang@163.com

Hui Yu
hui.yu@port.ac.uk

Yang Lu
luyanga@sina.com

Honghai Liu
honghai.Liu@port.ac.uk

1 Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Dong Nanhu Road 3888, Changchun, Jilin, China 130033

2 University of Portsmouth, Portsmouth, UK

3 Heilongjiang Bayi Agricultural University, Heilongjiang, China

challenging task to design an effective and efficient algorithm for robust face tracking. It becomes more difficult in cases such as a wide range of pose variation, partial occlusion, and complex background changing. These usually lead to tracking failure due to problems like template drift.

According to the investigation in literatures [11, 13], template drift is ascribed to the accumulation of small errors. It usually happens in template location when the template is updated each time. As representation of appearance feature extraction, frequent template renewal is required to keep the template up-to-date with the changing face appearance; at the same time, hasty update of the template will damage the integrity of the appearance feature. In order to obtain a good trade-off for the situation, template-updating strategies should be carefully designed.

Another problem is that many existing face detectors are often unable to cope with significant face appearance changes. Such challenges are particularly difficult when the algorithms heavily rely on human face feature extraction. A well-known face detector introduced by Viola and Jones [14] has been widely used over the years, which uses a boosted cascade of simple features. It is shown that the detection rate of this algorithm is quite high but it drops noticeably for profile, rotated, or occluded faces [9]. Although modifications have been introduced to address various face poses, e.g.,[10], these modifications increase its processing time to a great extent in a real world application.

In this paper, we design a novel and efficient fusion algorithm combining the Supervised Descend Method (SDM) and Compressive Tracking (CT) for robust face tracking. SDM is used to correct drifting error of CT continuously during the most frontal face tracking. But when the face turns aside widely enough, a tracking failure of SDM will be resulted in. Then a confidence score is introduced to trigger the shifting to CT. With features extracted by gradient integral, CT is adopted to keep the face region being tracked until SDM re-detects and tracks the face again.

The major contribution of this paper is a novel on-line fusion method which remarkably alleviates the drifting problem in robust face tracking. In order to improve efficiency and robustness, several difficulties have been overcome. First, the benefits of SDM in frontal face tracking is utilized adequately to correct the tracking error of CT in real-time. Second, by replacing Haar-like features [1] with gradient features, we are able to use CT to keep tracking robustly when the appearance changes completely different from face-like features. Third, the confidence score as a threshold allows us to shift the tracking strategy more flexibly for the best performance. In contrast to individual SDM and CT method, our fusion algorithm achieves reliable face tracking in real applications by making SDM and CT work together smoothly.

The remainder of the paper is organized as follows. Section 2 briefly reviews the related works, especially supervised descent method and compressive tracking method. In Section 3, our proposed method is described in details. Section 4 presents some experiments and results. Finally, we conclude this paper in Section 5.

2 Related work

To deal with these mentioned-above difficulties, many tracking algorithms [1, 6, 15, 17, 18, 21] have been proposed in recent years. Babenko et al. [1] introduce multiple instance learning into online tracking where samples are considered within positive and negative bags or sets. Wright et al. [15] propose a general classification algorithm for face recognition via sparse representation. In such approaches, two crucial aspects on face tracking, feature extraction and dimensionality reduction, has become more active [6, 12, 17, 20].

Both feature extraction and dimensionality reduction can be posed as solving a nonlinear optimization problem in computer vision. Although Newton's method is generally regarded as the most robust, fast and reliable optimization tool, two main drawbacks remain to be considered in the context of visual tracking: differentiability and computational cost. To address these issues, Li et al. [12] extend the ℓ_1-tracker by using the orthogonal matching pursuit algorithm for solving the optimization problems efficiently. They make use of the sparse signal recovery power of compressive sensing to significantly reduce the computational complexity. This algorithm shows that it is possible to accelerate the CS signal recovery procedure for tracking by randomly projecting the original features to a much lower dimensional space. Xiong et al. [17] further develop a Supervised Descent Method (SDM) for minimizing a Non-linear Least Squares (NLS) function, which shows perfect performance in accuracy and efficiency. During training, SDM learns a sequence of descent directions that minimizes the mean of NLS functions sampled at different points. In tracking, SDM minimizes the NLS objective using the learned descent directions without computing the Jacobian nor the Hessian.

Recently, several approaches have successfully applied on sparse representation of features for robust visual tracking [6, 18, 21, 22]. An important benefit of using sparse representation is its robustness to a wide range of feature appearance variations. T. Zhang et al. [22] propose a computationally efficient sparse and low-rank representation tracking method. They adopt a linear combination of object and background to represent samples features. This combination could be efficiently computed by solving a low-rank, sparse representation problem. Grabner et al. [6] introduce an online boosting algorithm to alleviate the drift problem in which only the samples in the first frame are labeled and all the other samples are unlabeled. This method is particularly well suited for scenarios where the object leaves the field of view completely, but it throws away a lot of useful information by not taking advantage of the problem domain (e.g., it is safe to assume small inter-frame motion). K. Zhang et al. [21] demonstrate that with an appearance model based on features extracted in the compressed domain, Compressive Tracking (CT) algorithm can be more efficient and effective than many existing trackers. CT accomplishes an efficient dimension compression via a sparse measurement matrix, which is also used for projection of both positive and negative samples. The best candidate is discriminated by a simple naive Bayes classifier learned online. R. Xu et al. [18] further improve CT method by replacing the rectangle filters with single pixels. They demonstrate that it is relatively redundant to convolving the intensity with multi-scale rectangle filters. The calculation of features is gained by directly projecting on the original image pixels with the sparse measurement matrix, which is not only simple but efficient in computation.

3 Problem formulation and fusion framework

3.1 Supervised descent method

Despite performing efficiently within most ranges of face pose during tracking, SDM suffers from tracking failure while the face turns near the angle of orthogonal to the view line (shown by Fig. 1). It needs to re-catch the face via other detectors, e.g., OpenCV (Viola-Jones face detector, which is much suitable for frontal face detection). Thus it becomes difficult to apply this tracker in robust tracking scenarios.

SDM can be divided into two stages: training and tracking. Given an image $d \in \Re^{m\times 1}$ of m pixels, $d\ (x) \in \Re^{p\times 1}$ indexes p landmarks in the image. h is a non-linear feature extraction

Fig. 1 The *yellow bounding box* denotes the ground truth, the *green* one denotes SDM tracking region. As the face turns aside, SDM could not detect any SIFT features of a face gradually, which results in tracking failure. Then SDM will search the whole image with OpenCV face detector until a frontal face is detected. Note that the face in frame 82, 86, 88, 99 could not be detected by the OpenCV face detector

function (e.g., SIFT) and $h(d(x)) \in \Re^{128p\times 1}$ in the case of extracting SIFT features. During training, an initial configuration of the landmarks (x_0) is provided, $f(x_0)$ could be defined as SIFT features function at x_0. Also, the correct landmarks are known, and referred as x_*, which corresponds to the optimization results of x_0. In this setting, face alignment can be framed as minimizing the following function over Δx

$$f(x_0+\Delta x) = \|h(d(x_0+\Delta x)) - \phi_*\|_2^2 \tag{1}$$

Where $\phi_* = h(d(x_*))$ represents the SIFT values in the manually labeled landmarks. In the training images, ϕ_* and Δx are known.

The training stage can be summarized as follows: SDM will learn a series of descent directions and re-scaling factors in a supervised manner. So that it produces a sequence of updates $x_{k+1}=x_k+\Delta x_k$ starting from x_0 that converges to x_* in the training data. The first updates of x would be given as a linear combination of feature vector ϕ_0 and a bias term b_0. R_0 is a projecting matrix referred as a descent direction.

$$\Delta x_1 = R_0\phi_0 + b_0 \tag{2}$$

$$x_k = x_{k-1} + R_{k-1}\phi_{k-1} + b_{k-1} \tag{3}$$

As illustrated in Fig. 2, at each step during training, a new dataset $\{\Delta x_k, \phi_k\}$ can be created by recursively applying the update rule in Eq. 3 with previously learned R_{k-1}; b_{k-1}. A new set of training data is generated by computing the new optimal parameter update $\Delta x_k^i = x_*^i - x_k^i$ and the new feature vector. $\phi_k - h(d_i(x_k^i))$. R_k and b_k can be learned from a new linear regressor in the new training set by minimizing

$$\arg\min_{R_k,b_k} \sum_{d_i} \sum_{x_*^{ki}} \left\| \Delta x_*^{ki} - R_k\phi_k^i - b_k \right\|^2 \tag{4}$$

To use this training data in tracking, SDM detects face in each frame with the learned generic directions and the initial configuration landmarks estimated from the previous frame. A confidence score is obtained to evaluate the performance.

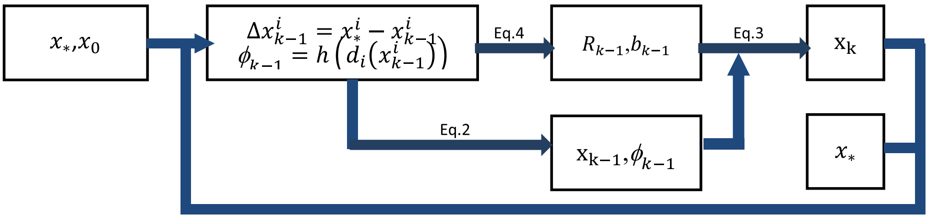

Fig. 2 Training loop of SDM

For tracking within most range of face pose, SDM shows excellent performance. However, while the face yaws to the orthogonal angle of the view line, there is only a profile face visible to the view. Or for the same pitch angle variation, the face is totally invisible. In such situation, it is impossible that SDM accomplish the detection of predicted landmarks with optimal displacement. There will be many errors caused by SIFT feature's incapability to distinguish between similar facial parts and other objects. A poor confidence score is generated to indicate the detecting result. While a tracking failure occurs inevitably, the method will re-detect face in full frame via the well-known OpenCV face detector [14], which is unreliable for profile detection and time consuming. But in the next several frames, the face appearance will not vary a lot due to a real-time frame rate. Hence, SDM will not re-track the face again until a similar frontal face appears.

3.2 Compressive tracking

Although the CT method has performed well in terms of efficiency, it still faces drifting problems, especially whilst a significant face appearance variation occurs (shown by Fig. 3). As a typical tracker based on appearance models, an error might be introduced with each update. These errors may accumulate and finally result in tracking failures in some situations. Looking at this problem from a classification point of view, we need to introduce a "supervisor" to calibrate the classifier.

Based on compressive sensing theories [3, 5], a compressible signal such as natural images could be reconstructed almost perfectly if it is compressed by a sparse random measurement matrix which satisfies the restricted isometric property (RIP) [2]. Therefore, CT uses this very sparse measurement matrix R to project the original image from a high dimensional feature space x to a low-dimensional compressed subspace v, as shown in Eq. 5. Generally R with entries is defined as shown in Eq. 6.

$$v = Rx \tag{5}$$

$$r_{ij} = \sqrt{s} \times \begin{cases} 1 & \textit{with probability } \frac{1}{2s} \\ 2 & \textit{with probability } 1-\frac{1}{s} \\ -1 & \textit{with probability } \frac{1}{2s} \end{cases} \tag{6}$$

For tracking, CT assumes that tracking window in the first frame has been determined. With the same sparse measurement matrix, CT projects some positive samples near the current target location and negative samples far away from it to update the classifier at each frame. To predict face location in the next frame, CT draws some samples around the current target location, and picks up the features of these samples under multi-scale filter around the face.

Fig. 3 The *yellow bounding box* denotes the ground truth, the *blue* one denotes CT tracking region. When Dudek moves his hand over his face, there will be a drastic changing in the tracking template appearance. Note that a drifting *error* occurs

Then a naive Bayes classifier is adopted to classify the object as shown in Eq. 7. The sample with the maximal classification score in $H(v)$ is the target of the next frame image.

$$H(v) = \log\left(\frac{\prod_{i-1}^{n} p(v_i|y=1)p(y=1)}{\prod_{i=1}^{n} p(v_i|y=0)P(y=0)}\right) = \sum_{i=1}^{n} \log\left(\frac{p\left(v_i|y=1\right.}{p(v_i|y=0)}\right) \tag{7}$$

According to our experiments on CT, although with accuracy and efficiency demonstrated successfully, drifting problems will emerge in the following cases. First, an inappropriate initial template is selected at the tracking outset. Second, a round face turning occurs. Third, a variation of template lasts for a certain period. As we know, to update the appearance model, a large quantity of positive samples and negative samples are computed, and a large amount of Haar-like features are employed, both of which are time consuming. To prevent the error which has been introduced to the tracker from accumulating severely, it is necessary to short the tracking loop and optimal the features selection. Therefore the initial tracking template and features selection will be crucial factors for CT robust tracking.

3.3 Fusion frame work

We propose a fusion method to deal with the template drifting problem. This fusion framework is to utilize the advantages of SDM and CT to compensate each other but avoid their drawbacks during tracking. To deal with the drifting problems with CT tracking, a more recent appearance model should be selected as the initial tracking template, which produces more reliable update parameters for the next few frames. Furthermore, a gradient integral feature has been shown to be more robust to scale and orientation change than a generalized Haar-like feature. On the other hand, in order to address the issue of re-catching in SDM, a local template covering the face region should be established as a re-detecting reference. This template should be tracked robustly by continuously updating appearances. The re-detecting reference could dramatically assist SDM to detect the face as soon as possible.

The fusion algorithm is illustrated in Algorithm 1. For each loop when SDM is tracking a face, it initializes a tracking template for CT in both scale and location, which means that the sparse measurement matrix and the naive Bayes classifier are updated by SDM in each loop. This ensures that a recent template for CT is prepared freshly. Hence the drifting error accumulation could be avoided to a large extent, and the features extracted from this instant template would represent the appearance model accurately so that the drifting probability is minimized. At the same time, a confidence score is given out to indicate the performance of SDM. Once the score is below a threshold, which represents that SDM is going to lose the face, CT is triggered to work. With the replacement of Haar-like feature by gradient integral, CT is adopted to keep the face template being tracked and updated. An advantage of data-independent of CT is utilized to improve robustness of tracking. For each loop, according to the tracking template provided by CT, SDM will try to detect and track the face. Once the confidence score is bigger than the threshold, SDM will return to dominate the tracking again. But when the score is still small, a larger tracking box will be generated for searching while CT keeps tracking and updating robustly.

Algorithm 1. Fusion tracking algorithm

Require: SDM classifier trained and CT sparse measurement matrix R available

Input: V_{t-1} is the template in previous frame.

1: while always do

```
Load new frame Xt
Vt = SDM(Vt−1)
score = SIFT similarity(Vt, Vt−1)
// To shift between SDM and CT based on the value of score
if score < Threshold then
Vt = RX s.t. X∈(Xt|Positive samples ∪Negative samples)
Vt = argmax(H(Vt)) via Eq. 7
end if
if score ≥ Threshold then
(λ,μ) = STDEV(Vt) // To initial CT by SDM tracking box for each loop
end if
end while
```

4 Experiments and discussion

We compare the performance of our fusion algorithm with the individual CT and SDM in terms of accuracy and robustness. In total, seven face sequences and their ground truth are selected for the experiments. All of these sequences are publicly available on the webpage [7]. In these sequences, face appearances have various conditions such as translation, rotation, scaling, illumination variation, etc. The proposed fusion method was implemented using C++ and the OpenCV library.

All the experiments are conducted under the setting described as below, First, the initial face region of CT is always from the ground truth rectangle R=[l; r; w; h], which can alleviate the drifting problem causing by inappropriate initial region selection. Second, the tracking result of SDM will be set to rectangle R=[0; 0; 0; 0] in case of SDM fails to track the face, which is illustrated in Fig. 5 by an abrupt ascension of the solid green curve. Third, since the scale of tracking window of CT is fixed during tracking whilst the one of SDM is adaptive, there will be some systematic errors existing.

We use the conventional metric center location error (CLE) [16] to verify the tracking accuracy. Generally, the tracking error is the Euclidean distance between the two centroids of the ground truth and the tracked region. These tracking boxes obtained from CT, SDM and Fusion algorithm are compared with the ground truth in the same sequence, respectively. In the following, we present both qualitative and quantitative evaluation of the proposed tracker, as well as compare it against CT and SDM methods. The experimental results are shown by frame snapshot and tracking error chart (Figs. 4 and 5). Different tracking methods and the Ground truth are color-coded, which the yellow one denotes Ground truth, blue for CT, green for SDM, and red for our fusion tracker.

4.1 Qualitative analysis

The face in the David sequence undergoes some pose, scale, illumination change and slightly occlusion. Figure 4a shows the tracking results at frame 71, 140, 302, and 415. As can be seen at frame 140, 414, SDM is suffering from an out of plane rotation of face and an abrupt illumination changes. At this moment, a lower confidence score is given to trigger fusion method to shift to CT tracking. In frame 302, CT causes drift from the target when David puts off and on his glasses. The fusion method can adopt SDM tracking result to correct the drift error and initialize the template for CT, then track the face robustly throughout the entire sequence, even though there is a little shift compared with ground truth in frame 414. These results show that the fusion algorithm compensates CT and SDM effectively for robust face tracking.

Fig. 4 Tracking results (*color-coded bounding boxes*) of three tracking methods. Ground truth - *yellow*. CT tracker - *blue*. SDM tracker - *green*. Our Fusion tracker - *red*

In the Dudek sequence, the tracked face is subject to changes in pose and appearance occlusion. The tracking results in frames 129, 208, 569, 1042 are shown in Fig. 4b. Note that in frame 208 the face is occluded by the moving hand. Our fusion algorithm keeps tracking whilst CT suffers from drifting and SDM loses the target. This occlusion affects the appearance

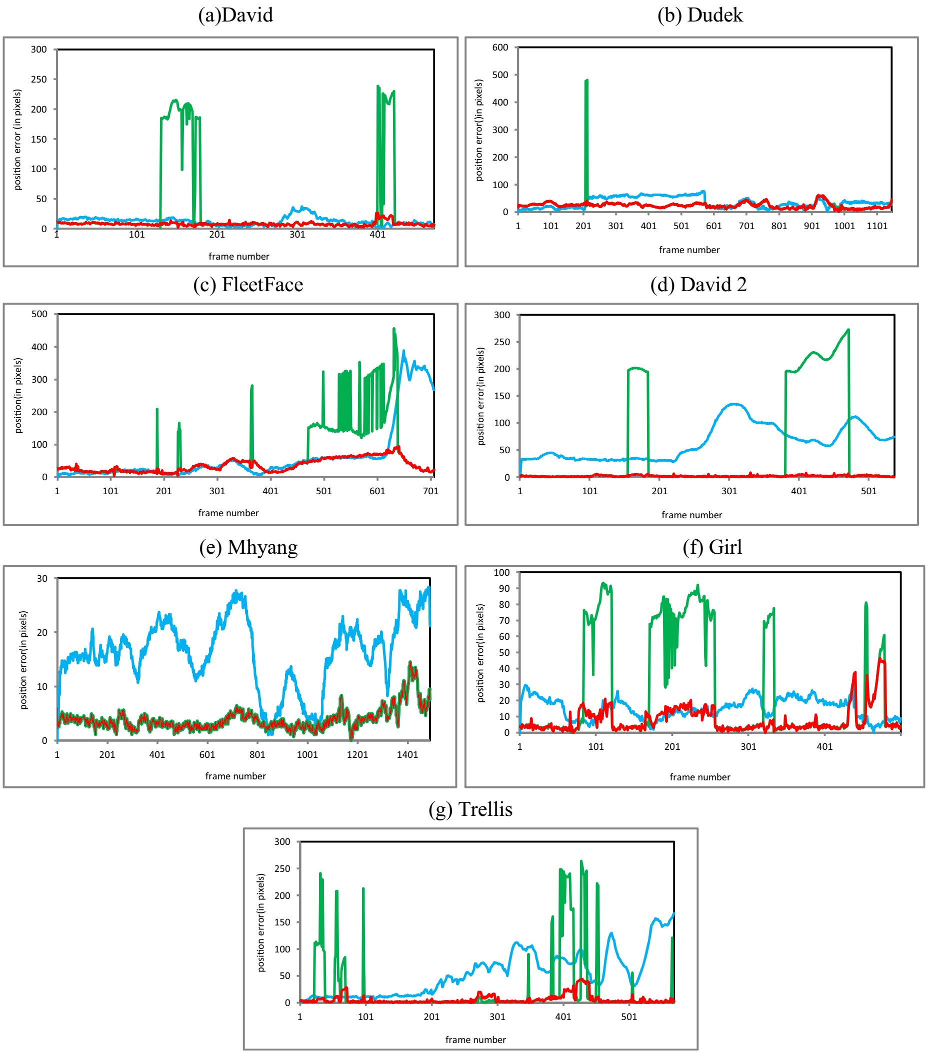

Fig. 5 Position *error* with respect to ground truth. Figures **a**–**g** show track-*box* position error w.r.t ground truth for different video sequences by means of different *colored curve*. CT tracker - *blue*. SDM tracker - *green*. Our Fusion tracker - *red*

feature extraction of the template which finally causes drifting problems to CT between frame 208 and frame 575. This makes the face fully invisible that result in SDM tracking failure temporarily. Once again our fusion tracker outperforms the other two trackers by tracking the moving face accurately throughout the sequence.

The Fleetface, David 2 and Mhyang sequences contain comprehensive face motion with significant translation, rotation, scale and background changes, which cause CT to drift and SDM to lose target respectively. Note that in Fig. 5e solid red overlaps solid green coincidently, which denote that fusion method is dominated by SDM during the whole tracking. The fusion algorithm works accurately and reliably as shown in Fig. 4c–e.

Results on the girl sequence are shown in Fig. 4f. Performance on this sequence exemplifies the accuracy and robustness of our fusion method to full occlusion, large pose variation and severe background clutter. Frame 100 shows complete occlusion of the girl's face as she swivels in the chair. A significant pose change is presented when the girl's face undergoes extensive 3D rotation about multiple axes. Additionally, this sequence includes multiple faces appearing around frame 495, which is similar in appearance to the girl's face. Nevertheless, our fusion tracker follows the girl face accurately and robustly, while CT starts drifting in frame 205, and SDM loses the target in frame 99.

The trellis sequence demonstrates drastic variations in illumination, pose, and background. The video is acquired in an outdoor environment where illuminations on both face and background change intricately. As shown in Fig. 4g, the cast shadow changes the appearance of the target face drastically when a person walks underneath a trellis covered by vines. Furthermore, the combined effects of pose and lighting variations along with a low frame rate make the tracking task extremely difficult. But thanks to the combination of benefits of CT and SDM, the fusion tracker overcomes problems and successfully tracks the face during the entire challenging sequence.

4.2 Quantitative analysis

Table 1 shows the mean and standard deviation of CLE. The Bold fonts indicate the best performance. It is noticeable that our fusion method outperforms the other two trackers in most of sequences in terms of mean and standard deviation.

Note that our algorithm achieves the second best result in the Mhyang sequence, with the difference of 0.001 pixels in mean error and 0.002 pixels in standard deviation from the first one. This slight difference is caused by the different initialization of tracking box just in the first frame. Note that in our fusion method, the score threshold is set as 0.55 regarding as the best fusing performance whilst 0.35 in SDM accordingly, which causes the tracking box differences between these two trackers.

Another second best result in standard deviation is given in the Girl sequence due to an abrupt extra face appearing in the scene at the end of the sequence, which confuses the tracker from following the right face within the interruption.

As shown in Fig. 5 a–g, we illustrate the track box position error w.r.t ground truth for different video sequences by means of different colored curve. It is obvious that the red curve (our tracker) outperforms the green one (SDM) and the blue one (CT) by means of a minimum error value. Note that the abrupt variation of the green one is caused by a temporary tracking

Table 1 Mean and Standard deviation of CLE (in pixels). The best results are shown in Bold fonts

Sequence	Mean			Standard deviation		
	CT	SDM	Fusion	CT	SDM	Fusion
David	12.495	33.711	**7.953**	6.763	66.828	**3.204**
Dudek	34.199	25.021	**23.599**	19.355	25.211	**9.549**
Fleetface	63.637	72.169	**36.440**	90.855	89.350	**20.542**
David2	63.669	50.006	**2.168**	31.941	90.470	**1.284**
Girl	14.886	25.267	**7.845**	**6.537**	32.726	8.370
Mhyang	16.271	**3.932**	3.934	6.289	**2.092**	2.093
Trellis	53.678	22.205	**5.740**	41.575	57.864	**8.511**

failure of SDM. Our tracker always achieves the best performance by keeping the lowest error value via fusing SDM and CT in all these sequences.

5 Conclusion

In this paper, we have presented a fusion algorithm as the human face tracker which significantly limits the drifting problem in real world applications. Using this fusion strategy, the benefits of two the-state-or-the-art methods are employed adequately while their drawbacks are overcome efficiently. We have kept the advantage of stability in frontal face tracking from the SDM method while avoided the drifting problems of CT. At the same time, by replacing Haar-like features with gradient features, we are able to use CT to keep robust tracking in situations where the appearance changes completely different from face-like features. We have tested the proposed fusion method using several real world video sequences containing various poses of facial appearance variations. The experimental results have demonstrated the effectiveness of the proposed algorithm.

Acknowledgments This work is partially supported by the State Scholarship Fund of China. Y. Lu is supported by the Natural Science Foundation of Heilongjiang Province of China under Grant F201428 and the 12th Five-Year-Plan in Key Science and Technology Research of agricultural bureau in Heilongjiang province of China under Grant HNK125B-04-03.

References

1. Babenko B, Yang MH, Belongie S (2011) Robust object tracking with online multiple instance learning. IEEE Trans Pattern Anal Mach Intell 33(8):1619–1632
2. Candes EJ, Tao T (2005) Decoding by linear programming. IEEE Trans Inf Theory 51(12):4203–4215
3. Candes EJ, Tao T (2006) Near-optimal signal recovery from random projections: universal encoding strategies? IEEE Trans Inf Theory 52(12):5406–5425
4. Demirkus M, Clark JJ, Arbel T (2014) Robust semi-automatic head pose labeling for real-world face video sequences. Multimed Tools Appl 70(1):495–523
5. Donoho DL (2006) Compressed sensing. IEEE Trans Inf Theory 52(4):1289–1306
6. Grabner H, Leistner C, Bischof H (2008) Semi-supervised on-line boosting for robust tracking. Comput Vis Eccv 5302:234–247,
   Pt I, Proceedings
7. https://sites.google.com/site/trackerbenchmark/benchmarks/v10
8. Hu CL et al (2014) An effective head pose estimation approach using Lie Algebrized Gaussians based face representation. Multimed Tools Appl 73(3):1863–1884
9. Huang C et al (2007) High-performance rotation invariant multiview face detection. IEEE Trans Pattern Anal Mach Intell 29(4):671–686
10. Jones MVP (2003) Fast multi-view face detection. Technical Report TR2003-96, Mitsubishi Electric Research Laboratories, Jul. 2003
11. Kaneko T, Hori O (2002) Template update criterion for template matching of image sequences. 16th International Conference on Pattern Recognition, Vol Ii, Proceedings, 1–5
12. Li HX, Shen CH, Shi QF (2011) Real-time visual tracking using compressive sensing. 2011 I.E. Conf Comput Vis Pattern Recogn (Cvpr), 1305–1312
13. Matthews I, Ishikawa T, Baker S (2004) The template update problem. IEEE Trans Pattern Anal Mach Intell 26(6):810–815
14. Viola P, Jones MJ (2004) Robust real-time face detection. Int J Comput Vis 57(2):137–154
15. Wright J et al (2009) Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 31(2):210–227
16. Wu Y, Lim J, Yang MH (2013) Online object tracking: a benchmark. 2013 I.E. Conf Comput Vis Pattern Recogn (Cvpr), 2411–2418

17. Xiong XH, De la Torre F (2013) Supervised descent method and its applications to face alignment. 2013 I.E. Conf Comput Vis Pattern Recogn (Cvpr), 532–539
18. Xu R, Gu XD, Wei BJ (2013) An improved real time compressive tracking. 2013 Second Iapr Asian Conference on Pattern Recognition (Acpr 2013), 2013:692–696
19. Yu H, Liu HH (2014) Regression-based facial expression optimization. IEEE Trans Human-Mach Syst 44(3):386–394
20. Zeisl B, et al (2010) On-line semi-supervised multiple-instance boosting. 2010 I.E. Conf Comput Vis Pattern Recogn (Cvpr), 1879–1886
21. Zhang KH, Zhang L, Yang MH (2012) Real-time compressive tracking. Comput Vis Eccv 7574:864–877, **Pt Iii**
22. Zhang TZ et al (2012) Low-rank sparse learning for robust visual tracking. Comput Vis Eccv 7577:470–484, **Pt Vi**

Xiaodong Jiang received the BSc and MSc degrees in Automatics from Jilin University in 1994 and in computer vision from Chinese Academy of Sciences in 1998, respectively. He was a research fellow in Applied Fundamental Research Division of Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. He is currently a visiting scholar in the School of Computing, University of Portsmouth, UK from 2014 to 2015. His research interests include computer graphics and image processing, weak and small target detection of real-time tracking.

Hui Yu is a Senior Lecturer with University of Portsmouth, UK. He received his PhD from Brunel University and previously held a research appointment with University of Glasgow, UK. Dr. Yu's research interests include computer vision, computer graphics and application of machine learning to these areas, particularly in Image/video processing and recognition, human computer interaction, human behavior understanding, robotics and geometric processing of human/facial performances.

Yang Lu received his MSc and PhD degrees from Northeast Petroleum University in Computer Application in 2005 and Oil and Gas Engineering in 2013, respectively. From December 2013 to December 2014, he was an academic visitor in the School of Computing, University of Portsmouth, UK. Now he is an associate professor at Heilongjiang Bayi Agricultural University. His research interests include machine learning and pattern recognition, computer vision and neural network.

Honghai Liu is a Professor in School of Computing as Head of the Intelligent Systems & Biomedical Robotics Group (ISR) at University of Portsmouth, UK. He received his PhD in Intelligent Robotics in 2003 from Kings College University of London, UK. He previously worked in industry on large-scale industrial control and system integration projects, and held research appointments at the Universities of London and Aberdeen. His research focuses on approximate computation, pattern recognition, multi-sensor based information fusion and analytics, human machine systems, advanced control, intelligent robotics and their practical applications, especially in cognition-driven bio-mechatronics and information abstraction, with an emphasis on approaches which could make contribution to the intelligent connection of perception to action using contextual information.