Journal of Information Hiding and Multimedia Signal Processing ©2018 ISSN 2073-4212
Ubiquitous International Volume **9**, Number **1**, January **2018**

Structural Local Sparse Tracking Method Based on Multi-feature Fusion and Fractional Differential

Wei-hua Niu and He-Jin Yuan
Department of Computer
North China Electric Power University
No.619, Younghua North Street, 071000, Baoding, China
tusiniuweihua@163.com; yhj_1977@163.com

Peng Zhao
State Grid Hebei Electric Power Company
No.32, Fuqiang Street, 050021, Shijiazhuang, China
zpnew@163.com

Received September, 2017; revised November, 2017

Abstract. *In view of incomplete using single feature to describe the objects of interest and effectiveness of local sparse representation model, we propose a novel structural local sparse tracker based on multi-feature fusion and fractional differential. Under the framework of particle filter, firstly, we partition target image region into nine different sub blocks. The gray feature and HOG feature of each sub block are extracted, and the two kinds of features are combined to express the sparse representation of the target. Secondly, using a kernelizable accelerated proximal gradient (KAPG) method, the sparse representation model can be solved efficiently. Lastly, sub blocks of different positions are viewed as different categories, we employ the block with the same category and the representation coefficients to reconstruct the block, likelihood function is constructed based on reconstruction error to determine the best candidate particle, which can achieve accurate tracking of the target. In tracking process, we employ sparse representation and Hamming distance based on fractional differential to update the dictionary templates adaptively. Compared with other tracking methods on benchmark classical image sequences, both qualitative and quantitative evaluations demonstrate that the proposed method is competitive to the state-of-the-art trackers on challenging benchmark video sequences with illumination changes, background clutter, occlusion, fast motion and deformation.*
Keywords: object tracking, sparse representation, multi-feature fusion, fractional differential, Hamming distance

1. **Introduction.** Visual tracking is an important research topic in the field of computer vision, which has been widely used in many fields, such as intelligent monitoring, vehicle navigation, human computer interaction and so on. In recent years, although many tracking methods has been proposed. But it is still a challenging problem. The reasons are mainly concentrated in some special cases (such as partial occlusion, light illumination variation, morphological changes and background clutter) of tracking accuracy, robustness and real-time.

Popular tracking method can be divided into two categories[1-8]: generative or discriminative approaches. Discriminative method regards the tracking problem as two valued classification problem, which is intended to separate target from background. It also uses information from both target and background. Generative method formulates tracking problem as searching for the most similar to target model, which is generally based on either templates or subspace models. In order to adapt to the change of target appearance caused by dynamic environment, the appearance model of target should keep dynamic update.

Recently, sparse representation has been used in computer vision successfully, including image enhancement, target recognition and visual tracking. Mei[9, 10] applied sparse representation to visual tracking, for the purpose to solve minimize problem, they took overall representation of target as appearance model. However, tracking is not stable enough when target appearance changes with high method complexity. Liu[11] proposed a tracking method based on local sparse appearance model, which employed sparse coefficient histogram and mean shift method to achieve the goal of tracking, but the method used static sparse dictionary, which would lead to tracking failure when there are similar objects with target object in the scene. Zhong[12] based on global template and local description, proposed sparsity-based discriminative classifier (SDC) and sparsity-based generative model (SGM), which SDC model is used to effectively distinguish between target and background and SGM model is used to the spatial information of blocks in order to deal with occlusion problem.

Xu[1] proposed a target tracking method based on structural local sparse appearance model (ASLA), which extract target gray level feature and construct a dictionary for overlapping of target area, then solve minimization paradigm l_1 with constructed over complete dictionary on target candidate, and target candidate with the minimum reconstruction error is used as tracking result. The method extracted a new vector from main feature of each block region with alignment-pooling technology, realized the dimension reduction and reduced computational complexity in some sense. However, the tracking method is relatively simple to describe the feature selection of target. Many researches have indicated the ability of different features to describe target is different, sharing and complementarity between different features can improve performance of individual features, and construction of local overlapping block dictionary method repeat calculate a lot of background information, which will affect real-time performance and robustness of tracking process with the more computing times of background information, the more likely it is to generate drift in the process of tracking.

In view of the above, inspired by the combination of sparse representation and successful application in visual recognition, we proposed a method of structural local sparse tracking based on multi-feature and fractional differential (MFLS). The main features of our method are: 1) using a new target board section block method; and 2) extracting multiple features (gray feature and HOG feature) to describe the target; and 3) using fractional differential and Hamming distance construct update weight sequence of adaptive update template. The method can overcome the situation of poor ability to describe target with single feature, give full play to the advantages of different feature description ability, improve tracking accuracy and tracking robustness, and reduce the possibility of tracking drift generated in tracking process.

2. **Multi-feature Joint Sparse Representation.**

2.1. **Target image region local block.**

There are many methods to block target image area. Xu et al[1] proposed that the size of local block is 16×16 pixels, step is 8 pixels, so for 32×32 pixels tracking region generate 9 overlapping sub blocks. In this paper, the target image region is divided into 9 different sub blocks. Suppose the target image region is 30×30 pixels, then the sub blocks are, sub block 1 and sub block 2 are 10×20 pixels, sub block 3 and sub block 4 are 20×10 pixels, sub block 5 is 10×10 pixels, sub block

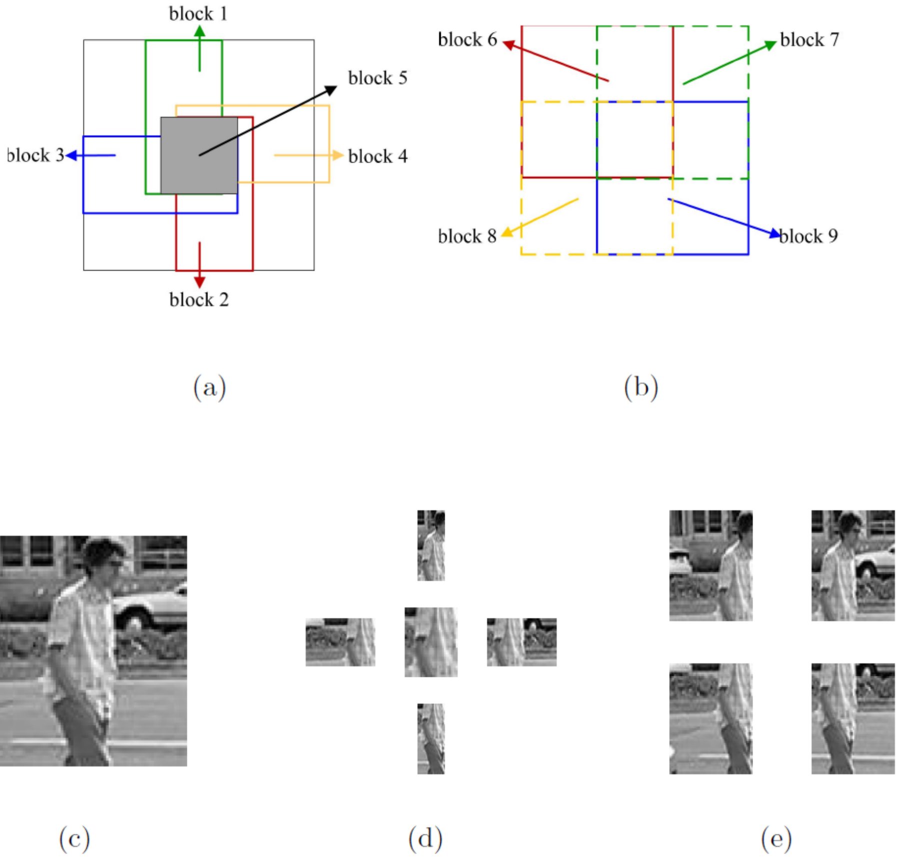


FIGURE 1. sub block 1 to sub block 5; (b) sub block 6 to sub block 9; (c) target image region; (d) image region of sub block 1 to sub block 5; (e) image region of sub block 6 to sub block 9

6, sub block 7, sub block 8, and sub block 9 are 20×20 pixels. As shown in Fig.1. The middle overlapping sub block 5 contains more information of target object, and multiple computations can enhance robustness of tracking.

2.2. **Feature Description of Target Local Block.** The classification idea is to take different position of the block as different categories, so block number of target image is equivalent to the number of categories. Suppose K different features, for each feature index $k = 1, 2, ..., K$, template dictionary is:

$$X^k = \left[X_1^k, X_2^k, \cdots, X_J^k\right] \in R^{d_k \times p} \tag{1}$$

where, J is the number of categories (number of local blocks divided by a single target template),$X_j^k \in R^{d_k \times p_j}$ is kth feature of jth local block, d_k is the dimension of kth feature.

In our method, the number of local blocks of each class is the same, because classification is according to the location of local blocks, local blocks of the same location belong to same class, so number of each class of local block is equal to number n of dictionary templates.

$$p = \sum_{j=1}^{J} p_j = n \times J \tag{2}$$

In the paper, we choose grayscale features with simple and intuitive advantages and HOG features with good invariance to deformation, so $K = 2$, but considering the generality of the method, we still use K as the number of features.

2.3. **Joint Sparse Representation.** The candidate sample (sampling particle) adopts same feature with the dictionary template, and each block is represented by all blocks in the dictionary with sparse representation. $Y_j^k = \left[\left[y_j^k\right]^{(1)}, \left[y_j^k\right]^{(2)}, \cdots, \left[y_j^k\right]^{(m)}\right] \in R^{d_k \times m}$ is the kth feature in m candidate samples of the jth class, then it can be linear represented with dictionary template block as:

$$Y_j^k = X^k w_j^k + \varepsilon_j^k, k = 1, 2, \cdots, K \tag{3}$$

where, $w_j^k \in R^{p\times m}$ is representation coefficient of kth feature in m candidate samples of the jth class, $\varepsilon_j^k \in R^{d_k \times m}$ is residual term. Take $w^k = \left[w_1^k, w_2^k, \cdots, w_J^k,\right] \in R^{p\times(m\times J))}$ as representation coefficient of kth feature, $w_j = \left[\left[w_j^1,\right]^T, \left[w_j^2,\right]^T, \cdots, \left[w_j^K\right]^T\right]^T \in R^{(K\times p)\times m}$ as representation coefficient of combining all K features corresponding to the jth class template. Therefore, multi-feature joint sparse representation can be formulated as solution of $l_{2,1}$ mixed norm regularization problem:

$$\min_{w_j} \frac{1}{2} \sum_{k=1}^{K} \left\|(Y_j^k - X^k w_j^k)\right\|_2^2 + \lambda \left\|w_j^T\right\|_{2,1} \tag{4}$$

where, $\left\|w_j^T\right\|_{2,1} = \sum_{i=1}^m \left\|\left[w_j^T\right]_i\right\|_2$, $\left[w_j^T\right]_i$ is w_j transposed the ith row.

Due to the different characteristic description of local block has different dimension, in order to solve the problem of inconsistency between features, therefore, it needs to deal dictionary template X^k and candidate sample Y_j^k with the following ways Y_j^k (kernelization):

$$X^k \in R^{d_k\times p} \overset{KP}{\to} \left[X^k\right]^T X^k \overset{KK}{\to} X^k \in R^{p\times p} \tag{5}$$

$$Y_j^k \in R^{d_k\times m} \overset{KP}{\to} \left[X^k\right]^T Y_j^k \overset{KK}{\to} Y_j^k \in R^{p\times m} \tag{6}$$

where KP denotes kernelization processing and KK denotes kernelized.

After kernelizable processing, for arbitrary feature k, there are $X^k \in R^{p\times p}$ and $Y_j^k \in R^{p\times m}$.

The kernelizable dimension of different features is only related to the number of local blocks. Hence, all features share the same dimension, which can better fuse arbitrary features.

For every feature, dictionary template and candidate sample are mapped from original feature space to another high dimension space RKHS with a nonlinear function ϕ^k. In RKHS, for given kernel function g^k, there is $\phi^k(x_i)^T\phi^k(x_j) = g^k(x_i, x_j)$, the Eq.7 is rewritten as:

$$\min_{w_j} \frac{1}{2} \sum_{k=1}^{K} \left\|(\phi^k(Y_j^k - \phi^k(X^k) w_j^k))\right\|_2^2 + \lambda \left\|w_j^T\right\|_{2,1} \tag{7}$$

Let $G^k = \phi^k(X^k)^T\phi^k(X^k)$ is the dictionary template kernel matrix corresponding to the kth feature, $H_j^k = \phi^k(X^k)^T\phi^k(Y_j^k)$ is the kernel vector corresponding to the jth candidate sample of the kth feature. A simple way to use the kernel matrix is to use G^k and H_j^k as new features directly, the Eq.8 is rewritten as:

$$\min_{w_j} \frac{1}{2}\sum_{k=1}^{K}\left\|(H_j^k - G^k w_j^k))\right\|_2^2 + \lambda\left\|w_j^T\right\|_{2,1} \tag{8}$$

Taking use of kernelizable accelerated proximal gradient (KAPG)[13], the corresponding sparse coefficients are obtained by solving the sparse problem Eq.8.

2.4. **Adaptive Template Update Based on R-L Fractional Differential.** In the process of tracking target movement, because the influence of environment background clutter and illumination changes, using the original definition of the target template and the target candidate template matching, the matching accuracy will be reduced and the tracking effect will be impacted [14]. However, if the dictionary template update is too frequent, the accumulation error will eventually lead to tracker drift.

(1)Template update principle

In most tracking methods, the early tracking results are relatively accurate, so they should be stored in the dictionary template for longer time. We hope the new template update slightly faster and the old template update slightly slower. Fractional differential can improve signal high frequency component and non-linearly preserve signal medium and low frequency components meanwhile. Riemann-Liouville(R-L) fractional differential definition is mainly used to calculate the analytical solution of some simple functions.

Assume that duration is $[a, x]$, then v order of R-L differentiation to $f(x)$ is defined as [15]:

$${}_a\mathrm{D}_x^v f(x) = \frac{d^v f(x)}{[d(x-a)]^v} = \frac{1}{\Gamma(-v)}\int_a^x (x-\tau)^{-v-1} f(\tau)d\tau (v<0) \tag{9}$$

where v is the differential number of ${}_a\mathrm{D}_x^v f(x)$, $\Gamma(\cdot)$ is Gamma function, $\Gamma(v) = \int_0^\infty e^{-t}t^{v-1}dt$. In Eq.9, we get the conclusion that if signal $f(x)$ is a constant, the corresponding v derivative is not zero. Therefore, some relevant numerical methods will not be used to contain integral derivative differential equations. Based on this, Jumarie G improved the definition of R-L fractional differential[16].

Definition 1: Assume that $R \to R$, $x \to f(x)$ denote a continuous function, h is a discrete constant and $h > 0$, let $FW(h)f(x) := f(x+h)$ (':=' means left item is defined by right item), then v $(0 < v < 1)$ order of R-L differential of $f(x)$ is defined as:

$$\triangle^v f(x) := (FW-1)^v f(x) = \sum_{k=0}^{\infty}(-1)^k \binom{v}{k} f[x+(v-k)h] \tag{10}$$

Limit of Eq.10 is:

$$f^{(v)}(x) := \lim_{h\to 0}\frac{\triangle^v[f(x)-f(0)]}{h^v} \tag{11}$$

Eq.11 is close to the standard derivative definition. While $0 < v < 1$ and $f(x)$ is constant, the v order of $f(x)$ is 0.

Definition 2: Assume that $R \to R$, $x \to f(x)$ denote a continuous function, then v $(0 < v < 1)$ order differential of $f(x)$ is:

$${}_a\mathrm{D}_x^v f(x) = (f^{(v-1)}(x))' = \frac{1}{\Gamma(1-v)}\frac{d}{dx}\int_a^x (x-\tau)^{-v}(f(\tau)-f(0))d\tau \tag{12}$$

where $f^{(v)} = (f^{(v-n)}(x))^{(n)}$, $n \le v < n+1, n \ge 1$.

Contrasting Eq.9 and Eq.12, we will find Eq.12 includes constant $f(0)$ but Eq.9 does not. Here, we call Eq.12 as improved R-L fractional order differential.

If the duration of signal function $f(x)$ is $x \in [a, x]$, dividing $[a, x]$ into N equal intervals, where $h = 1$. Let $a = 0$ and $\delta = x/N$, from Eq.10 we can get follows:

$$
\begin{aligned}
{}_a\mathrm{D}_x^v &= \frac{1}{\Gamma(1-v)}\frac{d}{dx}\int_0^x (x-\tau)^{-v}(f(\tau)-f(0))d\tau \\
&= \frac{1}{\Gamma(1-v)}\frac{d}{dx}\int_0^x \tau^{-v}(f(x-\tau)-f(0))d\tau \\
&= \frac{1}{\Gamma(2-v)}\int_0^x \frac{fx-\tau'}{\tau^{v-1}}d\tau - \frac{1}{\Gamma(2-v)} \times \frac{f(0)}{x^{v-1}} \\
&= \frac{1}{\Gamma(2-v)}\sum_{k=0}^{N-1}\int_{k\delta}^{k\delta+\delta}\frac{df(x-\tau)}{d\tau} \times \frac{1}{\tau^{v-1}}d\tau - \frac{1}{\Gamma(2-v)} \times \frac{f(0)}{N^{v-1}\delta^{v-1}}
\end{aligned}
\tag{13}
$$

For infinitesimal interval $[k\delta, k\delta+\delta]$, let $f_k = f(x-k\delta)$, then

$$
\frac{df(x-\tau)}{d\tau} \cong \frac{f(x-k\delta)-f(x-k\delta-\delta)}{\delta} = \frac{f_k - f_{k+1}}{\delta} \tag{14}
$$

Due to the minimum interval between discretization digital image pixels is 1, $\delta = 1$, so

$$
{}_a\mathrm{D}_x^v \approx \frac{1}{\Gamma(3-v)}\sum_{k=0}^{N-1}(f_k - f_{k+1})[(k+1)^{2-v} - k^{2-v}] - \frac{1}{\Gamma(3-v)} \times \frac{(2-v)f_N}{N^{v-1}} \tag{15}
$$

Therefore, fractional order differential can be approximate acted as convolution between signal $f(x)$ and the coefficient function A. We can calculate any value of coefficient function A by Eq.13:

$$
\begin{cases}
\alpha_1 = \dfrac{1}{\Gamma(3-v)} \\
\alpha_2 = \dfrac{2^{2-v}-2}{\Gamma(3-v)} \\
\alpha_3 = \dfrac{3^{2-v} - 2 \times 2^{2-v} + 1}{\Gamma(3-v)} \\
\cdots \\
\alpha_k = \dfrac{(k+1)^{2-v} - 2 \times k^{2-v} + (k-1)^{2-v}}{\Gamma(3-v)} \\
\cdots \\
\alpha_n = \dfrac{(n-1)^{2-v} - n^{2-v} - (2-v)n^{1-v}}{\Gamma(3-v)}
\end{cases}
\tag{16}
$$

So the fractional order differential sequence can be constructed as $A = (\alpha_0, \alpha_1, \cdots, \alpha_n)$, where n represents the number of templates in a template dictionary. Because of the difference of order v, the weight of the template weight sequence is also different. Partial calculation results are shown in Table 1.

From Table 1, we can find that the feature of this sequence is that the front is greater weight and the back is smaller weight. By this way, the old template will be updated slowly, and the new template will be updated quickly, which can reduce the template drift problem. In order to obtain better effect of update template, we choose $v = 0.5$ as the update weight sequence.

(2) Adaptive template update method

In information encoding, Hamming distance is used to denote different numbers of corresponding positions of two word with same length.

TABLE 1. Update weight sequence based on fractional differential with different v

v	update weight sequence									
0.2	0.5965	0.8841	0.7517	0.6911	0.6518	0.6230	0.6006	0.5823	0.5669	0.5536
0.3	0.6474	0.8086	0.6312	0.5561	0.5093	0.4760	0.4505	0.4300	0.4130	0.3987
0.4	0.6995	0.7215	0.5153	0.4350	0.3868	0.3534	0.3284	0.3086	0.2925	0.2790
0.5	0.7523	0.6232	0.4057	0.3281	0.2832	0.2530	0.2307	0.2135	0.1997	0.1882
0.6	0.8050	0.5144	0.3039	0.2354	0.1972	0.1722	0.1542	0.1405	0.1296	0.1208
0.7	0.8571	0.3962	0.2114	0.1567	0.1275	0.1088	0.0956	0.0858	0.0781	0.0719
0.8	0.9076	0.2699	0.1292	0.0917	0.0724	0.0604	0.0521	0.0460	0.0413	0.0376
0.9	0.9556	0.1372	0.0586	0.0398	0.0305	0.0248	0.0210	0.0183	0.0162	0.0146

Assume that $D_i, i = 1, 2, \cdots, n$, denotes template dictionary, which includes n templates, and T denotes object template, then we can obtain Hamming distance between every template in the dictionary and object template:

$$H_i = \sum_{j=1}^{K} D_i(j) \oplus T_j, j = 1, 2, \cdots, K \tag{17}$$

where K is the feature number of every template, $D_i(j)$ is the jth feature of D_i in the template dictionary, T_j is the jth feature of object template T, $\oplus$ is similarity XOR symbol. $H_i = 0$, denotes template D_i and object template T are completely consistent, which indicates two templates are very similar. $H_i = K$, denotes template D_i and object template T are totally different, which indicates two templates are widely different.

Based on fractional differential coefficient $\alpha_i, i = 1, 2, \cdots, n$, according update strategy that the first to enter the template dictionary is updated lower expectations, we can add weight to each template Hamming distance H_i and choose the biggest difference template to update. The formula is:

$$H' = max\left\{\frac{H_i}{\alpha_i}\right\} \tag{18}$$

The update method based on formula(18) is:

$$D_i' = \beta T + (1 - \beta)H' \tag{19}$$

where H' is current maximum difference template, T is object template, D_i' is the updated template, $\beta = \frac{H'}{K}$ is update parameter. The greater H', the greater difference between H' and T, the greater change of tracking target, the greater β which indicates update template demand is relatively larger. The smaller H', the smaller difference between H' and T, the smaller change of tracking target, the smaller β, which indicates update template demand is relatively less. So the integrated template update formula is:

$$D_i' = \frac{H'}{K} gT + (1 - \frac{H'}{K}) gH' \tag{20}$$

According to the Hamming distance between current template target H' and matching target template T , the template updating Eq.20 can be adjusted adaptively. The adaptive adjustment of template update can adapt to the change of the intensity or brightness of tracking target, and improve the accuracy and robustness of tracking.

3. **Tracking Method.** In this paper, the particle filter framework is implemented to track the target object and the method step is shown in Algorithm 1.
Algorithm 1 Tracking method

Structural local sparse tracking method based on multi-feature fusion and fractional differential
Input: Target initial state and target sequence image
Output: The target state of the current frame and updated dictionary template
1) Template initialization: tracking target location is obtained with manual way in the first frame, using Gauss distribution to generate m particles and constructing target template space according to KNN method to get the first 10 frames of tracking target.
2) Target area block: overlapping blocks of dictionary templates and candidate samples, extracting pixel gray value and HOG features of each block and carrying out kernelization processing.
3) Joint sparse representation: using all blocks in the dictionary template to complete multi-feature joint sparse representation of every block in candidate samples and solving sparse coefficients by KAPG.
4) Target state estimation: solving the reconstruction error of each sample particle, obtaining tracking results of the current frame according to the principle of minimum reconstruction error.
5) Template update: updating dictionary template using combined R-L fractional differential and Hamming distance method adaptively.
6) If the last frame has been reached, then tracking process is completed, or else jumping to 2).

4. **Experimental results and analysis.** In order to verify the effectiveness and the advancement of our method, we take use of 10 image sequences to finish tracking experiment, which includes challenging factors in tracking scenarios with illumination changes, background clutter, occlusion, shifting, etc. At the same time, we compared our method with five of state-of-the-art trackers on challenging benchmark [7], where multiple instance learning (MIL) [17] and Tracking-learning-detection (TLD) [18] are discriminative tracking methods, incremental visual tracker (IVT) [19], multi-task sparse tracker (MTT) [20] and adaptive structural local sparse tracker (ASLA) [1] are generative tracking methods. In experiments, our parameters are set to $m = 600$, regularization parameter $\lambda = 0.01$, dictionary template is updated every 5 frames. The challenging video sequences are: faceocc1, faceocc2, car4, carDark, david, trellis, skating1, singer1, bolt, deer. Experimental results are shown in Fig.2.

4.1. **Qualitative evaluation.** Various state-of-the-art trackers on challenging benchmark video sequences in Fig.2 with occlusion, illumination changes, deformation, background clutter and fast motion are analyzed as follows.
1) Occlusion
For the purpose of verification tracking performance in occlusion situation, we adopted serious occlusion image sequences faceocc1 andfaceocc2 as the tracking target. Fig.2(a) and Fig.2 (b) show that when serious occlusion occurs, the method can still accurately track the target. Compared with other 5 tracker, our method owned higher success rate, which shows that the method also owned better robustness in occlusion.
2) Illumination variation
Fig.2 (c), Fig.2 (d), Fig.2 (e), Fig.2 (f) and Fig.2 (h) show the tracking results of image sequences with obvious illumination variation such as car4, david, trellis and singer1. We

FIGURE 2. Partial tracking results

can see that the tracking effect of proposed MFLS method and ALSA method are better than other 4 tracking methods.

3) Deformation

Fig.2 shows the tracking results of image sequences with deformation such as avid, trellis, skating1 and bolt. At the same time, these sequences have other challenging factor, such as david and trellis contain illumination variation, skating1and bolt contain occlusion. Regardless of the tracking results or the average success rate, our tracking method is significantly better than the other 5 tracking method.

4) Background clutter

Fig.2 (d) and Fig.2 (j) show the tracking results of image sequence with background clutter such as carDark and deer. At the same time, these sequences have other challenging factor, such as carDark contains illumination variation, deer contains fast motion, etc. The proposed method in this paper achieves a higher tracking success rate mean value in tracking the two sequences.

5) Fast motion

In addition to background clutter, the image sequence deer contains the main challenges of fast motion and motion blur. This challenge can still continue to tracking target stably.

4.2. **Quantitative evaluation.** In this paper, we take tracking success (overlap) rate curve and average value of the success rate as index of quantitative evaluation. Assume that the target true rectangle region is r_a and tracking results rectangle region is r_t respectively, the overlap score is defined as $score = area(r_t \cap r_a)/area(r_t \cup r_a)$. If the overlap score of tracking method on a frame image is larger than that of the overlapped threshold t_0, it is considered that the tracking method is successful on this frame image. According to the definition of tracking method in an image sequence, tracking success rate is the ratio between successful tracking image frames and total image sequences frames.

The comparison of average tracking success rate between our method and 5 different tracking methods in 10 image sequences is shown in Table 2.

TABLE 2. Average success rate of tracking

Video	IVT	MTT	ASLA	MIL	TLD	MFLS
faceocc1	0.83	0.83	0.31	0.57	0.52	0.89
faceocc2	0.79	0.74	0.71	0.73	0.56	0.80
carDark	0.51	0.59	0.77	0.22	0.32	0.81
car4	0.74	0.54	0.87	0.29	0.58	0.92
bolt	0.02	0.03	0.15	0.51	0.17	0.73
deer	0.21	0.58	0.59	0.20	0.39	0.72
david	0.37	0.52	0.45	0.43	0.59	0.69
trellis	0.39	0.60	0.62	0.35	0.21	0.76
skating1	0.08	0.51	0.80	0.21	0.08	0.81
singer1	0.49	0.84	0.72	0.42	0.41	0.84

From Table 2, we could find that our method has the highest average success rate in all compared methods, which indicate our tracker is the most accurate. For further verify performance of our method, we compared 51 image sequences in benchmark with success rate curve of temporal robustness evaluation (TRE). In Fig.3, (a) is the comparison of overall performance, (b) is the comparison of illumination variation, (c) is the comparison of occlusion, (d) is the comparison of fast motion, (e) is the comparison of background clutter, (f) is the comparison of deformation.

From Fig.3(a), we can conclude that our method is better in the overall performance compared with other tracking methods. And from Fig.3(b)- Fig.3(h), we can find the tracking performance of our method is relatively better in challenging situation such as illumination variation, occlusion, fast motion, background clutter, deformation.

5. **Conclusion.** In this paper, we proposed a new method based on multi-feature joint sparse representation theory and fractional differential. At the same time, using the multiple features of target and the target tracking region local block, it can not only make up for the disadvantage of single gray feature describe target model, but also can deal with occlusion in the image sequence. Our method adopted pixel gray value feature and HOG feature in the description of target feature. Actually, target of the different attributes image sequences, such as serious occlusion, illumination change, fast motion, deformation, background clutter, motion blur, etc., has its own description of features. According to the attribute characteristic of image sequence, choosing suitable characteristic and adjusting confidence weights of different features adaptively is the question which needs to solve in the future.

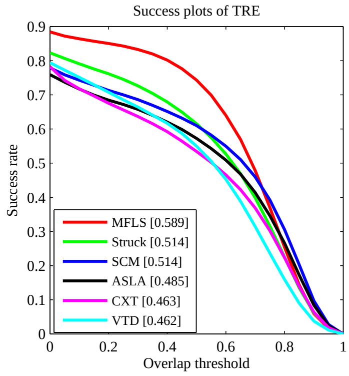
Competing Interests. The authors declare that they have no competing interests.

Acknowledgment. The work reported in this paper was supported by "the Fundamental Research Funds for the Central Universities (2017MS156)". We also acknowledge the anonymous reviewers for comments that lead to clarification of the paper.

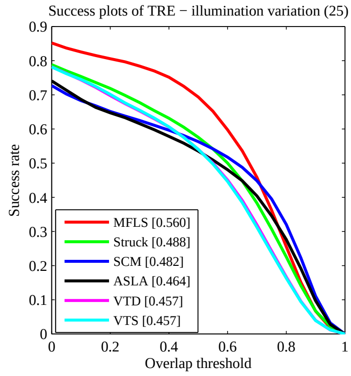
REFERENCES

[1] J. Xu, H. C. Lu, M. Yang, Visual tracking via adaptive structural local sparse appearance model. *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1822-1829, 2012.

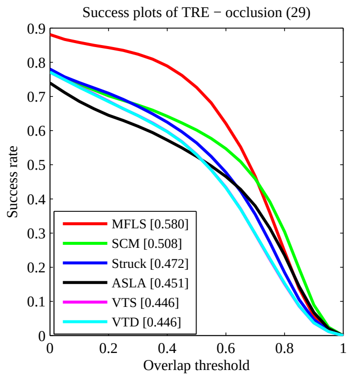
[2] H. A. Abdelali, F. Essannouni, L. Essannouni, et al, A new moving object tracking method using particle filter and probability product kernel. *IEEE Conference on Intelligent Systems and Computer Vision*, pp.1-6, 2015.



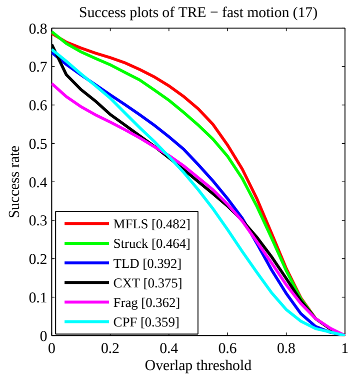
(a) overall performance



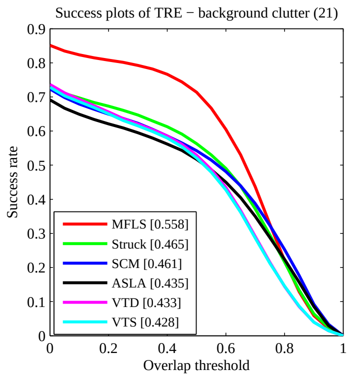
(b) illumination variation



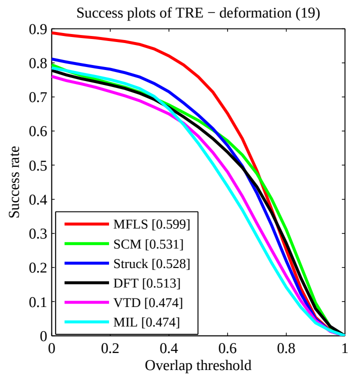
(c) occlusion



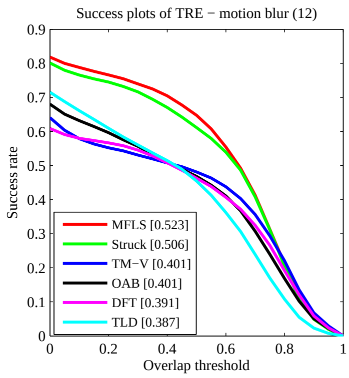
(d) fast motion



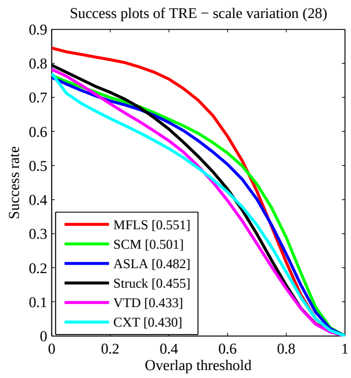
(e) background clutter



(f) deformation



(g) motion blur



(h) scale variation

Figure 3. The plots of TRE with attributes using the success rate metric

[3] J. Kajiki, M. Ochi, T. Ikeya, et al, Moving object tracking apparatus, and moving object tracking method. *Astronomy & Astrophysics*, vol.327, no.1, pp.90-106, 2015.

[4] M. L. Gao, J. Shen, L. J. Yin, A novel visual tracking method using bat algorithm. *Neurocomputing*, vol.177, no.C, pp.512-619, 2016.

[5] G. B. Lee, A fast moving object tracking method by the combination of covariance matrix and kalman filter algorithm. *Journal of the Korea Institute of Information and Communication Engineering*, vol.19, no.6, pp.1477-1484, 2015.

[6] X. T. Yuan, S. S. Yan, Visual classification with multi-task joint sparse representation. *IEEE Conference on Computer Vision and Pattern Recognition*, pp.3493-3500, 2012.

[7] Y. Wu, J. Lim, M. Yang, Online object tracking: A benchmark. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2013.

[8] P. Kyung, H. Jae, S. Dong, Tracking method of adaptive for inserting an artificial Lens. *International Journal of Modeling & Optimization*, vol.5, no.1, pp.6-9, 2015.

[9] X. Mei, H. B. Lin, Robust visual tracking and vehicle classification via sparse representation. *IEEE Transactions on Pattem Analysis and Machine Intelligence*, vol.33, no.13, pp.2259-2272,2011.

[10] X. Mei, H. B. Lin, Robust visual tracking using l_1 minimization. *IEEE 12th International Conference on Computer Vision*, pp.1436-1443,2009.

[11] B. Y. Liu, J. Z. Huang, L. Yang, et al, Robust visual tracking with local sparse appearance model and k-selection. *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-8, 2011.

[12] W. Zhong, H. C. Lu, M. Yang, Robust object tracking via sparsity-based collaborative model. *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1838-1845, 2012.

[13] X. T. Yuan, X. B. Liu, S. S. Yan, Visual classification with multi-task joint sparse representation. *IEEE Transactions on Image Processing*, vol.21, no.10, pp. 4349-4360, 2012.

[14] W. Hu, A new meanshift robust tracking method for fast-moving target. *Computer Applications and Software*, vol.31, no.1, pp.140-143, 2014.

[15] K. B. Oldham, J. Spanier, The fractional calculus: theory and applications of differentiation and integration to arbitrary order. *London: Academic Press*, 1974.

[16] G. Jumarie, Table of some basic fractional calculus formulae derived from a modified Riemann-Liouville derivative for non-differentiable functions. *Journal of Applied Mathematics Letters*, vol.22, no.3, pp.378-385, 2009.

[17] B. Babenko, M. Yang, S. Belongie, Visual tracking with online multiple instance learning. *IEEE Conference on Computer Vision and Pattern Recognition*, pp.983-990, 2009.

[18] Z. Kalal, J. Matas, K. Mikolajczyk, P-N learning: bootstrapping binary classifiers by structural constraints. *IEEE Conference on Computer Vision and Pattern Recognition*, pp.49-56, 2010.

[19] D. A. Ross, J. Lim, R. Lin, et al, Incremental learning for robust visual tracking. *International Journal of Computer Vision*, vol.77, no.1, pp.125-141, 2008.

[20] T. Z. Zhang, B. Ghanem, S. Liu, et al, Robust visual tracking via multi-task sparse learning. *IEEE Conference on Computer Vision and Pattern Recognition*, pp.2042-2049, 2012.

[21] Y. Xie, W. S. Zhang, C. H. Li, et al, Discriminative object tracking via sparse representation and online dictionary learning. *IEEE Transactions on Cybernetics*, vol.44, no.4, pp.539-553, 2014.

[22] Y. Q. Wang, X. F. Yan, W. Zhang, An Algorithm of Feather and Down Target Detection and Tracking Method Based on Sparse Representation. *International Conference on Measuring Technology and Mechatronics Automation*, pp.87-90, 2015.

[23] Z. P. Zhou, M. Z. Zhou, J. Li, Object tracking method based on hybrid particle filter and sparse representation. *Multimedia Tools & Applications*, vol.76, no.2, pp.1-15, 2016.

[24] G. F. Wang, X. Y. Qin, F. Zhong, et al, Visual Tracking via Sparse and Local Linear Coding. *IEEE Transactions on Image Processing*, vol.24, no.11, pp.3796-3809, 2015.

[25] Z. J. Han, J. B. Jiao, B. C. Zhang, et al, Visual object tracking via sample-based adaptive sparse representation (AdaSR). *Pattern Recognition*, vol.44, no.9, pp.2170-2183, 2011.