Incremental Robust Nonnegative Matrix Factorization for Object Tracking

Fanghui Liu[1], Mingna Liu[2,3], Tao Zhou[3], Yu Qiao[1], and Jie Yang[1](✉)

[1] Institute of Image Processing and Pattern Recognition,
Shanghai Jiao Tong University, Shanghai, China
lfhsgre@gmail.com, {zhou.tao,qiaoyu,jieyang}@sjtu.edu.cn

[2] Shanghai Institute of Spaceflight Control Technology, Shanghai, China
mingnal@126.com

[3] Infrared Detection Technology Research and Development Center,
China Aerospace Science and Technology Corporation, Shanghai, China

Abstract. Nonnegative Matrix Factorization (NMF) has received considerable attention in visual tracking. However noises and outliers are not tackled well due to Frobenius norm in NMF's objective function. To address this issue, in this paper, NMF with $L_{2,1}$ norm loss function (robust NMF) is introduced into appearance modelling in visual tracking. Compared to standard NMF, robust NMF not only handles noises and outliers but also provides sparsity property. In our visual tracking framework, basis matrix from robust NMF is used for appearance modelling with additional ℓ_1 constraint on reconstruction error. The corresponding iterative algorithm is proposed to solve this problem. To strengthen its practicality in visual tracking, multiplicative update rules in incremental learning for robust NMF are proposed for model update. Experiments on the benchmark show that the proposed method achieves favorable performance compared with other state-of-the-art methods.

Keywords: Incremental robust NMF · Appearance model · Visual tracking

1 Introduction

Appearance modelling is an overriding concern in visual tracking and has been studied for several years [12]. One widespread adoption of appearance modelling is with generative method [11,17], which aims to search the most similar candidate to the target with minimizing reconstruction error. The representative generative methods include but are not limited to subspace learning [8], sparse representation [5,16].

Recently, Nonnegative Matrix Factorization (NMF) based on subspace learning has been successfully used in visual tracking with variety works [13,15]. Wang et al. [9] utilize projective NMF for appearance modelling. Liu *et al.* [7] propose an inverse coding view for visual tracking, in which NMF is served as a feature coder. NMF seeks for two nonnegative matrices $\mathbf{U}$ (the basis matrix) and $\mathbf{V}$

A. Hirose et al. (Eds.): ICONIP 2016, Part II, LNCS 9948, pp. 611–619, 2016.
DOI: 10.1007/978-3-319-46672-9_68

(the coefficient matrix) to represent the original data matrix $\mathbf{X}$ ($\mathbf{X} \approx \mathbf{UV}$). The basis matrix $\mathbf{U}$ is a representation to the target in a low-dimensional space, like PCA. The objective function in standard NMF uses the least square error function which is well known but not stable. Because real data may contain many undesirable noises and outliers due to partial occlusions in visual tracking. Though these above methods can deal with outliers by imposing additive various sparse constraints on $\mathbf{U}$ or $\mathbf{V}$, they could not tackle this problem in essence, and suffer from significant performance degradation. To address this issue, data reconstruction function is formulated in the form of robust matrix norms such as ℓ_1 norm or $L_{2,1}$ norm [6,14]. Robust NMF (RNMF) not only handles outliers and noises but also incurs almost the same computation cost as standard NMF.

Motivated by the above issues, this paper introduces RNMF into visual tracking framework. Specially, incremental learning for RNMF should be taken into consideration for online tracking. The main novelties of the proposed method include: (a) We present an iterative algorithm including the standard nonnegative least square method for target coefficients and the soft-threshold method to obtain sparse error coefficients. (b) Multiplicative update rules for incremental RNMF are provided, as the indispensable parts of online visual tracking. (c) Model update scheme (First-in and First-out) for visual tracking is proposed with incremental RNMF learning. Experiments on different videos illustrate our tracker can handle partial occlusion and other challenging factors.

2 RNMF and Its Incremental Learning

2.1 Review: Robust NMF with $L_{2,1}$ norm

Given a data matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N] \in \mathbb{R}^{M \times N}$, robust NMF is defined as,

$$\begin{aligned} &\min_{\mathbf{U},\mathbf{V}} \|\mathbf{X} - \mathbf{UV}\|_{2,1} \\ &s.t. \ \ \mathbf{U} \geq 0, \ \ \mathbf{V} \geq 0 \end{aligned} \tag{1}$$

where the basis matrix $\mathbf{U} \in \mathbb{R}^{M \times K}$, and $L_{2,1}$ norm is defined as $\|\mathbf{X}\|_{2,1} = \sum_{i=1}^{N} \sqrt{\sum_{j=1}^{M} x_{ij}^2} = \sum_{i=1}^{N} \|\mathbf{x}_i\|$ and $\mathbf{x}_i$ is the i-th column of $\mathbf{X}$. Because the objective function is not convex on both $\mathbf{U}$ and $\mathbf{V}$, and $L_{2,1}$ norm is harder to solve than standard NMF with Frobenius norm. The corresponding objective function can be rewritten as in [6],

$$\mathcal{O}(\mathbf{U}, \mathbf{V}) = Tr[(\mathbf{X} - \mathbf{UV})\mathbf{D}(\mathbf{X} - \mathbf{UV})^\top] \tag{2}$$

where $\mathbf{D}$ is a diagonal matrix and its diagonal element is given by $D_{ii} = 1/\|\mathbf{x}_i - \mathbf{U}\mathbf{v}_i\|$. The iteratively updating algorithm, proposed by Kong $etal.$ [6], obtains a local minimum value of robust NMF. The iteration formula is following:

$$u_{jk}^{t+1} \leftarrow u_{jk}^{t} \frac{(\mathbf{XDV}^\top)_{jk}}{(\mathbf{UVDV}^\top)_{jk}}, v_{ki}^{t+1} \leftarrow v_{ki}^{t} \frac{(\mathbf{U}^\top\mathbf{XD})_{ki}}{(\mathbf{U}^\top\mathbf{UVD})_{ki}} \tag{3}$$

2.2 Incremental Learning for Robust NMF

The implicit assumption behind the standard incremental NMF [1] is that the previous coefficient matrix $\mathbf{V}$ has no effect on the incremental process when a new sample $\mathbf{x}$ is added: $[\mathbf{X}, \mathbf{x}] \approx \mathbf{U} \times [\mathbf{V}, \mathbf{v}]$. Similarly, in our incremental updating for RNMF, define that $\mathbf{X}_{t+1} = [\mathbf{X}_t, \mathbf{x}]$, $\mathbf{V}_{t+1} = [\mathbf{V}_t, \mathbf{v}]$, $\mathbf{U}_{t+1}$, and $\mathbf{D}_{t+1} = diag(\mathbf{D}_t, d_{t+1})$ are the corresponding matrices when the $(t+1)$-th sample arrives. The corresponding objective function is formulating as,

$$\begin{aligned}
&\mathcal{O}(\mathbf{U}_{t+1}, \mathbf{V}_{t+1}) = \mathcal{O}(\mathbf{U}_{t+1}, \mathbf{v}) \\
&= Tr[(\mathbf{X}_{t+1} - \mathbf{U}_{t+1}\mathbf{V}_{t+1})\mathbf{D}_{t+1}(\mathbf{X}_{t+1} - \mathbf{U}_{t+1}\mathbf{V}_{t+1})^\top] \\
&= Tr(\mathbf{X}_{t+1}\mathbf{D}_{t+1}\mathbf{X}_{t+1}^\top) - 2Tr(\mathbf{X}_{t+1}\mathbf{D}_{t+1}\mathbf{V}_{t+1}^\top\mathbf{U}_{t+1}^\top) \\
&\quad + Tr(\mathbf{U}_{t+1}\mathbf{V}_{t+1}\mathbf{D}_{t+1}\mathbf{V}_{t+1}^\top\mathbf{U}_{t+1}^\top)
\end{aligned} \tag{4}$$

We expend $\mathbf{D}_{t+1}$ and $\mathbf{V}_{t+1}$ to separate $d_{t+1} = 1/\|\mathbf{x} - \mathbf{U}_{t+1}\mathbf{v}\|$ and $\mathbf{v}$ for convenience,

$$\begin{aligned}
\mathcal{O}(\mathbf{U}_{t+1}, \mathbf{v}) &= Tr(\mathbf{X}_t\mathbf{D}_t\mathbf{X}_t^\top) + d_{t+1}Tr(\mathbf{x}\mathbf{x}^\top) \\
&\quad - 2Tr(\mathbf{X}_t\mathbf{D}_t\mathbf{V}_t^\top\mathbf{U}_{t+1}^\top) - 2d_{t+1}Tr(\mathbf{x}\mathbf{v}^\top\mathbf{U}_{t+1}^\top) \\
&\quad + Tr(\mathbf{U}_{t+1}\mathbf{V}_t\mathbf{D}_t\mathbf{V}_t^\top\mathbf{U}_{t+1}^\top) + d_{t+1}Tr(\mathbf{U}_{t+1}\mathbf{v}\mathbf{v}^\top\mathbf{U}_{t+1}^\top)
\end{aligned} \tag{5}$$

Incremental Update Rules on $\mathbf{U}_{t+1}$: Given $\mathbf{V}_{t+1}$, let ψ_{ij} be a Langrange Multiplier for the nonnegative constraint on $\mathbf{U}_{t+1}$. The relevant Langrange function corresponding to Eq. (5): $\mathcal{L}_{\mathbf{U}_{t+1}} = \mathcal{O}(\mathbf{U}_{t+1}, \mathbf{v}) + Tr(\psi\mathbf{U}_{t+1})$. The partial derivatives of $\mathcal{L}_{\mathbf{U}}$ with respect to $\mathbf{U}_{t+1}$ is:

$$\begin{aligned}
\frac{\partial \mathcal{L}_{\mathbf{U}}}{\partial \mathbf{U}_{t+1}} &= \mathbf{x}^\top\mathbf{x}\frac{\partial d_{t+1}}{\partial \mathbf{U}_{t+1}} - 2\mathbf{X}_t\mathbf{D}_t\mathbf{V}_t^\top - 2Tr(\mathbf{x}\mathbf{v}^\top\mathbf{U}_{t+1}^\top)\frac{\partial d_{t+1}}{\partial \mathbf{U}_{t+1}} \\
&\quad - 2d_{t+1}\mathbf{x}\mathbf{v}^\top + 2\mathbf{U}_{t+1}\mathbf{V}_t\mathbf{D}_t\mathbf{V}_t^\top + 2d_{t+1}\mathbf{U}_{t+1}\mathbf{v}\mathbf{v}^\top \\
&\quad + \mathbf{v}^\top\mathbf{U}_{t+1}^\top\mathbf{U}_{t+1}\mathbf{v}\frac{\partial d_{t+1}}{\partial \mathbf{U}_{t+1}} + \psi
\end{aligned} \tag{6}$$

where $\frac{\partial d_{t+1}}{\partial \mathbf{U}_{t+1}} = d_{t+1}^3(\mathbf{x} - \mathbf{U}_{t+1}\mathbf{v})\mathbf{v}^\top$. Using the KKT condition for each elements in ψ and $\mathbf{U}_{t+1}$ to satisfy $\psi_{ij}u_{ij} = 0$, we get the equations for u_{ij}. And then these equations lead to the following update rule shown on $\mathbf{U}_{t+1}$ in Eq. (8), where $S_1 = 2d_{t+1}^3Tr(\mathbf{x}\mathbf{v}^\top\mathbf{U}_{t+1}^\top)$, $S_2 = d_{t+1}^3\mathbf{v}^\top\mathbf{U}_{t+1}^\top\mathbf{U}_{t+1}\mathbf{v}$ and $S_3 = \mathbf{U}_{t+1}\mathbf{v}\mathbf{v}^\top$.

Incremental Update Rules on $\mathbf{v}$: Let ϕ_j be a Langrange Multiplier for the nonnegative constraint on $\mathbf{v}$. The relevant Langrange function with respect to Eq. (5): $\mathcal{L}_{\mathbf{v}} = \mathcal{O}(\mathbf{U}_{t+1}, \mathbf{v}) + Tr(\phi\mathbf{v})$. The partial derivatives of $\mathcal{L}_{\mathbf{v}}$ with respect to $\mathbf{v}$ is:

$$\begin{aligned}
\frac{\partial \mathcal{L}_{\mathbf{v}}}{\partial \mathbf{v}} &= \mathbf{x}^\top\mathbf{x}\frac{\partial d_{t+1}}{\partial \mathbf{v}} - 2d_{t+1}\mathbf{U}_{t+1}^\top\mathbf{x} - 2Tr(\mathbf{x}\mathbf{v}^\top\mathbf{U}_{t+1}^\top)\frac{\partial d_{t+1}}{\partial \mathbf{v}} \\
&\quad + 2d_{t+1}\mathbf{U}_{t+1}^\top\mathbf{U}_{t+1}\mathbf{v} + \mathbf{v}^\top\mathbf{U}_{t+1}^\top\mathbf{U}_{t+1}\mathbf{v}\frac{\partial d_{t+1}}{\partial \mathbf{v}} + \phi
\end{aligned} \tag{7}$$

where $\frac{\partial d_{t+1}}{\partial \mathbf{v}} = d_{t+1}^3 \mathbf{U}_{t+1}^\top (\mathbf{x} - \mathbf{U}_{t+1}\mathbf{v})$. Using the KKT condition for each element in $\mathbf{v}$ with $\phi_j v_j = 0$, we get the following update rule for $\mathbf{v}$ shown in Eq. (9), where $S_4 = \mathbf{U}_{t+1}^\top \mathbf{U}_{t+1}\mathbf{v}$. Therein, we give relatively detailed description about incremental learning for robust NMF, which can be used in the proposed visual tracking method effectively.

3 Proposed Robust NMF Tracker

3.1 Particle Filter in Visual Tracking Framework

Particle filter has been widely applied to visual tracking for state estimation. The implicit rationale behind particle filter is to estimate the posterior distribution

$$(\mathbf{U}_{t+1})_{ij} \leftarrow (\mathbf{U}_{t+1})_{ij} \cdot \frac{\{2\mathbf{X}_t\mathbf{D}_t\mathbf{V}_t^\top + d_{t+1}^3\|\mathbf{x}\|^2\mathbf{U}_{t+1}\mathbf{v}\mathbf{v}^\top + S_1\mathbf{x}\mathbf{v}^\top + 2d_{t+1}\mathbf{x}\mathbf{v}^\top + S_2S_3\}_{ij}}{\{d_{t+1}^3\|\mathbf{x}\|^2\mathbf{x}\mathbf{v}^\top + S_1S_3 + 2\mathbf{U}_{t+1}\mathbf{V}_t\mathbf{D}_t\mathbf{V}_t^\top + 2d_{t+1}S_3 + S_2\mathbf{x}\mathbf{v}^\top\}_{ij}} \tag{8}$$

$$v_i \leftarrow v_i \cdot \frac{\{d_{t+1}^3\|\mathbf{x}\|^2 S_4 + S_1\mathbf{U}_{t+1}^\top\mathbf{x} + 2d_{t+1}\mathbf{U}_{t+1}^\top\mathbf{x} + S_2S_4\}_i}{\{d_{t+1}^3\|\mathbf{x}\|^2\mathbf{U}_{t+1}^\top\mathbf{x} + S_1S_4 + S_2\mathbf{U}_{t+1}^\top\mathbf{x} + 2d_{t+1}S_4\}_i} \tag{9}$$

$p(\mathbf{x}_t|\mathbf{Y}_{1:t})$ approximately by a finite set of random sampling particles. Given some observed image patches at t-th frame $\mathbf{Y}_{1:t} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{t-1}\}$, the state of the target $\mathbf{x}_t$ can be estimated recursively.

$$p(\mathbf{x}_t|\mathbf{Y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t) \int p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}|\mathbf{Y}_{1:t-1}) \mathrm{d}\mathbf{x} \tag{10}$$

where $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ denotes state transition between two consecutive frames, and $p(\mathbf{y}_t|\mathbf{x}_t)$ denotes observation model that estimates the likelihood of observing $\mathbf{y}_t$ at state $\mathbf{x}_t$. The optimal state at t-th frame is obtained by maximizing the approximate posterior probability:

$$\mathbf{x}_t^* = \underset{\mathbf{x}_t}{\operatorname{argmax}}\, p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{x}_{t-1}) \tag{11}$$

3.2 Observation Model

The observation model in our proposed tracking method is shown in Fig. 1. At the first m frames, target templates are collected to be as the initial data matrix $\mathbf{X}$. RNMF decomposes it into the basis matrix $\mathbf{U}$ composed of several basis vectors. Compared to standard NMF, the learned basis matrix $\mathbf{U}$ not only provides sparse representation property but also retains the Euclidean property of data vectors attribute to $L_{2,1}$ norm. Motivated by [10], the estimated target can be modeled by a linear combination of basis vectors and trivial templates.

$$\mathbf{y} = \mathbf{U}\mathbf{z} + \mathbf{e} = \begin{bmatrix} \mathbf{U} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{z} \\ \mathbf{e} \end{bmatrix} \tag{12}$$

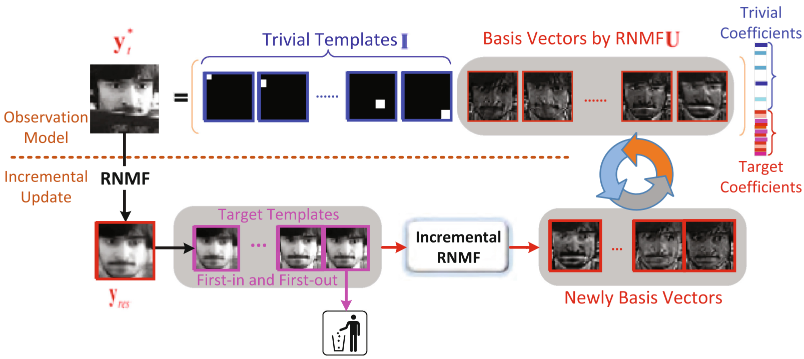

Fig. 1. Illustration of the proposed method by RNMF

where $\mathbf{y}$ denotes an observation vector, $\mathbf{z}$ indicates the coefficients of basis vectors, and $\mathbf{e}$ can be assumed to sparse error. The objective function in our proposed method is formed as,

$$\min_{\mathbf{z},\mathbf{e}} \frac{1}{2}\|\mathbf{y}-\mathbf{U}\mathbf{z}-\mathbf{e}\|_2^2+\lambda\|\mathbf{e}\|_1 \quad s.t. \;\; \mathbf{z}\geq 0 \tag{13}$$

Obviously, there is no closed-form solution for the optimization problem with Eq. (13). The iterative algorithm to compute $\mathbf{z}$ and $\mathbf{e}$ is proposed.

When the optimal $\hat{\mathbf{e}}$ is obtained, the optimization problem inverts to $\mathcal{O}(\mathbf{z}) = \frac{1}{2}\|\hat{\mathbf{y}}-\mathbf{U}\mathbf{z}\|_2^2$, $s.t.$ $\mathbf{z}\geq 0$, where $\hat{\mathbf{y}}=\mathbf{y}-\hat{\mathbf{e}}$. It is a typical nonnegative least square problem and easily solved with several iterations. When the optimal $\hat{\mathbf{z}}$ is obtained, Eq. (13) is equal to $\mathcal{O}(\mathbf{e}) = \frac{1}{2}\|\bar{\mathbf{y}}-\mathbf{e}\|_2^2+\|\mathbf{e}\|_1$, where $\bar{\mathbf{y}}=\mathbf{y}-\mathbf{U}\hat{\mathbf{z}}$. It has a closed-form solution by soft-threshold operation which is defined as $S_\lambda(x) = \mathrm{sgn}(x)(|x|-\lambda)$. The algorithm for solving $\mathbf{z}$ and $\mathbf{e}$ is shown in **Algorithm** 1. When target coefficients $\mathbf{z}_t$ obtained at t-frame, the observation likelihood can be measured by the reconstruction error of each observed image patch.

$$p(\mathbf{y}_t^i|\mathbf{x}_t^i) \propto \exp(-\|\mathbf{y}_t^i-\mathbf{U}\mathbf{z}_t^i\|_2^2) \tag{14}$$

Algorithm 1. Algorithm for $\hat{\mathbf{z}}$ and $\hat{\mathbf{e}}$.

Input: An observation vector $\mathbf{y}$, RNMF basis matrix $\mathbf{U}$, the regularization parameter is $\lambda = 0.05$

Output: The optimal $\hat{\mathbf{z}}$ and $\hat{\mathbf{e}}$

Set: stop error ε.

Initialize $i = 0$ and $\mathbf{e}$ with random positive values.

Repeat

  Update $\mathbf{z}^{i+1} = \min_{\mathbf{z}} \frac{1}{2}\|\hat{\mathbf{y}}-\mathbf{U}\mathbf{z}\|_2^2$, $s.t.$ $\mathbf{z}\geq 0$;

  Update $\mathbf{e}^{i+1} = S_\lambda(\bar{\mathbf{y}}-\mathbf{U}\mathbf{z}^{i+1})$;

  $i := i+1$;

Until $\frac{\|\mathbf{z}^{i+1}-\mathbf{z}^i\|_2}{\|\mathbf{z}^i\|_2} \leq \varepsilon$;

3.3 Online Model Update

Along with a new tracking result obtained, appearance model of the target needs to be updated. The computational complexity of NMF is proportional to the number of samples. Therefore, we retain the number of target templates by a "First-in First-out" scheme shown in Fig. 1. The promptly update target templates, as the new data matrix $\mathbf{X}_{t+1}$, are sent to incremental RNMF model. The previous basis matrix is substituted for the recalculated basis matrix by Eq. (8).

It is worth noting that the optimal candidate $\mathbf{y}^*$ can be not added into the target template set directly. Just because the optimal candidate may contain noises and outliers due to partial occlusions. If the observed vector (the optimal candidate $\mathbf{y}^*$) is used for model updating, our basis vectors would be contaminated by occlusions (the book) shown in Fig. 1.

Considering robust NMF can effectively handle noises and outliers, RNMF is used to address this issue. At t-th frame, the corresponding encoding coefficient vector $\mathbf{v}_t^*$ is utilized to represent the reconstruction sample $\mathbf{y}_{res} = \mathbf{U}_t \mathbf{v}_t^*$ with the positive templates $\mathbf{U}_t$ of RNMF. It is added into the target template set to obtain the newly basis vectors by incremental learning for RNMF. For the tradeoff between the computation and the model fitness, the basis matrix $\mathbf{U}$ is recalculated every five frames in our test.

4 Experiments

In this section, we test the proposed model on Object Tracking Benchmark (OTB) [12] with 29 trackers and 51 sequences. Besides, CN [2], SST [16] and KCF [4] are taken into comparison.

Setup: The proposed tracker was implemented in MATLAB on a PC with Intel Xeon E5506 CPU (2.13 GHz) and 24 GB memory. The following parameters were used for our tests: each patch of image was normalized to 32×32 pixels; the number of initial target templates was $N = 50$ respectively from the first 5 frames; the number of basis vectors was set to 16; kNN was fixed to 5.

We show the overall performance of OPE for our tracker and compare it with some other state-of-the-arts (ranked within top 10) as shown in Fig. 2(a), (b).

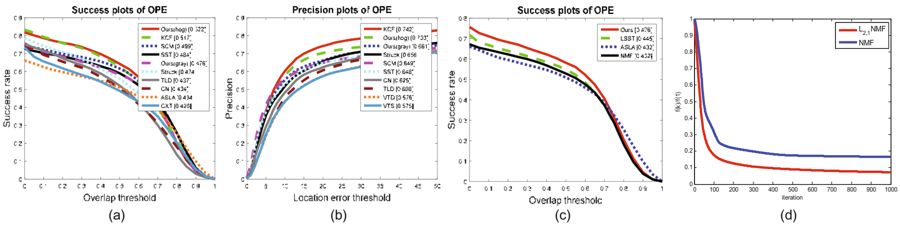

Fig. 2. (a,b): Plots of OPE. The performance score for each tracker is shown in the legend. For each figure, the top 10 trackers are presented for clarity; (c): Success plot of OPE for self-validation; (d) Convergence analysis for RNMF and NMF

Our method with HOG feature ranks the first on the success plot and the second on precision plot followed by KCF tracker. The proposed tracker with raw feature shows a comparable performance.

Specially, in Fig. 2(c), ASLA, LSST and NMF method are also provided just for self-validation. In NMF tracker, basis vectors are obtained by NMF and updated by incremental NMF. Results on OPE demonstrate that our tracker with robust NMF is with improvements than the conventional NMF tracker. Compared to LSST with PCA vectors, basis vectors obtained by robust NMF have a higher representation ability to capture the object's appearance.

Results: Figure 3 shows screen shots of tracking results from different trackers including Struck [3], ASLA [5], SCM [17], SST [16] and KCF [4].

Fig. 3. Representative frames of some sampled tracking results. And subfigures from left to right in the first row are *David3*, *Sylvester*. In the second row, these subfigures are *Basketball* and *Freeman1*. In the second row, these subfigures are *Freeman1*, *Carscale* and *Pedestrian*.

Occlusion: Occlusion is the most important attributes which lead to tracking drifts. In *David3* sequence, when David passes through the tree, SST and Struck suffer from severe drifts. ASLA and SST lose tracking accuracy to some extent. In *Basketball* sequence, SST, LSST and the proposed method accurately track the basketball player when the player is occluded by others. Results on *Occlusion* attribute show that our method can effectively handle outliers due to occlusions.

Deformation: Trackers based on NMF have natural advantage on this attribute because of part-based representation. Many methods including SST, Struck and ASLA are not adapt to appearance changes in *David3*, *Sylvester* and *Basketball* sequences. They show difficult in capturing appearance changes, where the appearance becomes much dissimilar to their initial one. Our method shows more robust and performs well without drifts.

Other attributes: There are others remaining attributes we do not refer to. The *Freeman1* sequence contains *out-of-plane rotation* and the challenging issues in

Carscale are *occlusion* and *scale variation*. The *Deer* sequence is accompanied with *fast motion* and *motion blur*. The proposed method not only locates the target to obtain accurate appearance model but also covers the target accurately.

Besides, our tracker also performs well in an infrared video *Pedestrian* than other trackers, such as KCF, SCM and SST. In summary, our test results on these videos have shown that the proposed tracker are effective and robust to heavy occlusions and deformation.

Convergence analysis: We compare the convergence properties (i.e. speed, objective function value) between RNMF and NMF on these sequences shown in Fig. 2(d). Compared with the standard NMF with Frobenius norm, RNMF with $L_{2,1}$ norm has better precision and is relatively faster to reach steady state. However, the computation time cost on RNMF is higher than NMF due to the updating rule on the diagonal matrix $\mathbf{D}$ in each iteration. It is worth mentioning that computational complexity of these two methods is both $\mathcal{O}(tMNK)$ when the multiplicative update is supposed to stop after t iterations.

5 Conclusion

This paper proposes an incremental learning for robust NMF method in visual tracking. Multiplicative update rules are derived with convergence verification. RNMF and its incremental update rules are incorporated into visual tracking framework. The optimal problem can be solved by an iterative algorithm with nonnegative least square method and soft-threshold operator. Quantitative and qualitative comparisons with other state-of-the-art methods on OTB have demonstrated the effectiveness and robustness of the proposed tracker.

Acknowledgment. This research is partly supported by NSFC, China (No: 6127 3258), 863 Plan, China (No. 2015AA042308), USCAST2015-10 and USCAST2013-07.

References

1. Bucak, S.S., Gunsel, B.: Incremental subspace learning via non-negative matrix factorization. Pattern Recognit. **42**(5), 788–797 (2009)
2. Danelljan, M., Khan, F.S., Felsberg, M., Weijer, J.V.D.: Adaptive color attributes for real-time visual tracking. In: IEEE Conference on Computer Vision and Pattern Recognition (2014)
3. Hare, S., Saffari, A., Torr, P.H.S.: Struck: structured output tracking with kernels. In: IEEE International Conference on Computer Vision, pp. 263–270 (2011)
4. Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: High-speed tracking with kernelized correlation filters. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 583–596 (2015)
5. Jia, X., Lu, H., Yang, M.H.: Visual tracking via adaptive structural local sparse appearance model. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1822–1829 (2012)

6. Kong, D., Ding, C., Huang, H.: Robust nonnegative matrix factorization using l21-norm. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, pp. 673–682. ACM (2011)
7. Liu, F., Zhou, T., Yang, J., Gu, I.Y.: Robust visual tracking via inverse nonnegative matrix factorization. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE (2016)
8. Ma, L., Zhang, X., Hu, W., Xing, J., Lu, J., Zhou, J.: Local subspace collaborative tracking. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4301–4309 (2015)
9. Wang, D., Lu, H.: On-line learning parts-based representation via incremental orthogonal projective non-negative matrix factorization. Signal Process. **93**(6), 1608–1623 (2013)
10. Wang, D., Lu, H., Yang, M.H.: Online object tracking with sparse prototypes. IEEE Trans. Image Process. **22**(1), 314–325 (2013)
11. Wang, N., Wang, J., Yeung, D.Y.: Online robust non-negative dictionary learning for visual tracking. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 314–325 (2013)
12. Wu, Y., Lim, J., Yang, M.H.: Object tracking benchmark. IEEE Trans. Pattern Anal. Mach. Intell. **38**(4), 1–1 (2015)
13. Wu, Y., Shen, B., Ling, H.: Visual tracking via online nonnegative matrix factorization. IEEE Trans. Circ. Syst. Video Technol. **24**(3), 374–383 (2014)
14. Zhang, H., Zha, Z.J., Yang, Y., Yan, S., Chua, T.S.: Robust (semi) nonnegative graph embedding. IEEE Trans. Image Process. **23**(7), 2996–3012 (2014)
15. Zhang, H., Hu, S., Zhang, X., Luo, L.: Visual tracking via constrained incremental non-negative matrix factorization. IEEE Signal Process. Lett. **22**(9), 1350–1353 (2015)
16. Zhang, T., Liu, S., Xu, C., Yan, S., Ghanem, B., Ahuja, N., Yang, M.H.: Structural sparse tracking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 150–158 (2015)
17. Zhong, W., Lu, H., Yang, M.H.: Robust object tracking via sparse collaborative appearance model. IEEE Trans. Image Process. **23**, 2356–2368 (2014)