

Research Article

Low-Rank Representation-Based Object Tracking Using Multitask Feature Learning with Joint Sparsity

Hyuncheol Kim and Joonki Paik

Department of Image, Chung-Ang University, Seoul 156-756, Republic of Korea

Correspondence should be addressed to Joonki Paik; paikj@cau.ac.kr

Received 13 August 2014; Revised 7 November 2014; Accepted 7 November 2014; Published 23 November 2014

Academic Editor: Sergei V. Pereverzyev

Copyright © 2014 H. Kim and J. Paik. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We address object tracking problem as a multitask feature learning process based on low-rank representation of features with joint sparsity. We first select features with low-rank representation within a number of initial frames to obtain subspace basis. Next, the features represented by the low-rank and sparse property are learned using a modified joint sparsity-based multitask feature learning framework. Both the features and sparse errors are then optimally updated using a novel incremental alternating direction method. The low-rank minimization problem for learning multitask features can be achieved by a few sequences of efficient closed form update process. Since the proposed method attempts to perform the feature learning problem in both multitask and low-rank manner, it can not only reduce the dimension but also improve the tracking performance without drift. Experimental results demonstrate that the proposed method outperforms existing state-of-the-art tracking methods for tracking objects in challenging image sequences.

1. Introduction

Object tracking is one of the well-known problems in computer vision with many applications including intelligent surveillance, human-computer interface, and motion analysis. In spite of significant success, designing a robust object tracking algorithm remains still challenging issue due to factors from real-world scenarios such as severe occlusion, scale and illumination variations, background clutter, rotations, and fast motions.

An appearance model-based tracking method, which evaluates the likelihood of an observed image patch belonging to object class, considers some critical factors such as object representation and representation scheme. The object representation can be categorized by adopted features [1, 2] and description models [3, 4]. The representation scheme can be either generative or discriminative. The generative methods regard the appearance modeling as finding the image observation with minimal reconstruction error [5, 6]. On the other hand, the discriminative methods focus on determining a decision region that distinguishes the object from the background [7, 8].

Various object tracking methods based on object appearance models can handle only moderate changes and usually fail to track when the object appearance significantly changes. As a result, an appearance model learning process is required for robust object tracking under challenging issues such as object deformation.

Recently, sparse representation-based ℓ_1 norm minimization methods have been successfully employed for object tracking [9–13], where an object is represented as one of multiple candidates in the form of sparse linear combination of a dictionary that can be updated to maintain the optimal object appearance model. Although the sparse representation-based method can robustly track an object with partial occlusion, the computational cost for the ℓ_1 norm minimization in each frame is still expensive.

Bao et al. applied the accelerated proximal gradient (APG) [14] approach to efficiently solve the ℓ_1 minimization problems for object tracking. Structured multitask tracking (MTT) was proposed by mining object discriminative structure between different particles rather than individually learning each particle [15]. Zhang et al. address their tracking performance as a fast solution of the MTT problem using

Bao's learning method of joint particle representation. By decomposing the sparse coefficients into two matrices for maintaining inliers and outliers, the robust MTT- (RMTT-) based object tracking was proposed in [16].

The low-rank sparse tracking (LRST) was proposed by representing all samples using only a few templates [17]. The LRST is improved with incremental learning method of low-rank features [18] and adaptive pruning with exploiting temporal consistency [19].

In spite of these improvements in MTT and LRST in the particle filter framework, the computational cost increases with the number of particles. Furthermore, MTT regards the sparse representations of sampled particles as independent state data without considering relationships between particles. For this reason, MTT may not work if the particles are drawn from specific probability distribution. The LRST-based methods are difficult to apply directly for object tracking in online video processing due to its structural computation complexity such as nuclear norm minimization.

To solve the above mentioned problems, we propose a novel object tracking algorithm based on multitask feature learning using joint sparsity and low-rank representation. We assume that the object representation can be incrementally optimized in the robust principal component analysis (RPCA) framework. The RPCA can be performed by decomposing the observations as the sum of a low-rank matrix and a sparse matrix; thus it can successfully recover the intrinsic subspace structure from corrupted observations. We extract features with low-rank representation within a few frames. After obtaining the subspace basis of object features, the features represented by all possible low-rank and the sparse property are learned using a variant of multitask feature learning framework. Finally, a novel incremental alternating direction method- (ADM-) based low-rank optimization strategy is efficiently applied for update of sparse error and features. The low-rank optimization problem for learning multitask features can be achieved by a few sequences of efficient closed form updating operation for the optimal state variables of object tracking.

2. Low-Rank Representation of Object with Joint Sparsity

2.1. Low-Rank Representation of Object. RPCA [20] was recently introduced to recover low-rank features in learning for the reduced subspace. Given a sparse matrix \mathbf{Y} with corrupted observation, RPCA attempts to exactly recover the low-rank matrix \mathbf{Z} and minimize sparse error \mathbf{E} . Let $\|\mathbf{Z}\|_* = \sum_i \sigma_i(\mathbf{Z})$ denote the nuclear norm of the low-rank matrix \mathbf{Z} such as the sum of singular values of \mathbf{Z} . Under this assumption, the principal component pursuit (PCP) is defined as in the following minimization:

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}} \quad & \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{1,1}, \\ \text{s.t.} \quad & \mathbf{Y} = \mathbf{Z} + \mathbf{E}, \end{aligned} \quad (1)$$

where $\|\cdot\|_{1,1}$ represents the $\ell_{1,1}$ norm and λ is a regularization parameter.

The assumption under RPCA framework [20], for disentangling the low-rank and sparse component, is that \mathbf{Z} has singular vectors and is not sparse. The singular value decomposition of \mathbf{Z} can be expressed as

$$\mathbf{Z} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top = \sum_{i=1}^N \sigma_i u_i v_i^\top, \quad (2)$$

where N is the rank of \mathbf{Z} , $\sigma_1, \dots, \sigma_N$ are positive singular values, and $\mathbf{U} = [u_1, \dots, u_N]$ and $\mathbf{V} = [v_1, \dots, v_N]$ are the matrices of left- and right-singular projection vectors.

The object tracking problem can be solved under the RPCA framework. Suppose we have n sampled particles and the correspondingly observed candidate of an object $\mathbf{Y} = [y_1, \dots, y_n]$ at time t . Each observation can be represented as a linear combination of m basis vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$, which spans the data space such as $\mathbf{Y} = \mathbf{U}_t \mathbf{W}_t$, where $\mathbf{U}_t = [\mathbf{u}_1, \dots, \mathbf{u}_m]$, and each column of $\mathbf{W}_t = [\mathbf{w}_1, \dots, \mathbf{w}_n]$ indicates the discriminative feature representation of each particle observation. In this paper, we utilize the strength of both subspace learning and sparse representation for object appearance model in the principal component analysis (PCA) as [21]

$$\mathbf{Y} = \mathbf{U}\mathbf{W} + \mathbf{E}. \quad (3)$$

In feature learning literature, the multitask feature learning [22, 23] employs the $\ell_{2,1}$ norm minimization to involve features selection across tasks under a strict assumption that all tasks share a common underlying representation.

The goal of the proposed object tracking framework is to search particles that have the most similar feature to previous tracking result. Since particles are densely sampled around the current object state, in order to estimate the optimal low-rank features in iterative steps, we focus on all possible nonsmooth sparse errors, joint sparsity of features based on low-rank, and $\ell_{1,1}$ and $\ell_{1,2}$ constraints that produce solution for the within- and between-feature-tasks sparsity. With this consideration, we formulate the following object tracking problem by adopting three equality constraints as

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{E}} \quad & \lambda_1 \|\mathbf{W}\|_* + \lambda_2 \|\mathbf{W}\|_{1,2} + \lambda_3 \|\mathbf{W}\|_{1,1} + \lambda_4 \|\mathbf{E}\|_{1,1} \\ \text{s.t.} \quad & \mathbf{Y} = \mathbf{U}\mathbf{W} + \mathbf{E}, \end{aligned} \quad (4)$$

where $\|\mathbf{C}\|_{p,q} = (\sum_j (\sum_i |c_{ij}|^p)^{q/p})^{1/q}$ is called p, q norm of matrix \mathbf{C} and $\lambda_1, \lambda_2, \lambda_3$, and λ_4 are regularization parameters.

In the proposed object tracking formulation in (4), we should minimize the rank of the matrix for features of all object candidates. Since there is no closed form solution for the rank minimization problem, we replace the rank minimizing of matrix with its convex envelope by the nuclear norm as $\|\mathbf{W}\|_*$. Moreover, mining similarities between all particle structures can improve the tracking results. However, in the object tracking process, some candidates are completely different from others when particles are sampled from an abnormally large region. In order to address this problem, we place the matrix \mathbf{W} in shared feature $\|\mathbf{W}\|_{1,2}$ and nonshared sparse feature $\|\mathbf{W}\|_{1,1}$. The shared inlier feature exploits the similarities of particles while the nonshared

sparse feature considers possible appearance variations of object and background between candidates, and only a small number of candidates are required to reliably represent the observation of each candidate. In order to reduce sparse errors, we minimize the $\ell_{1,1}$ norm of error matrix \mathbf{E} . This error assumption has been originally introduced in [9] in the presence of object occlusion.

In order to solve this objective function in (4), we introduce three equality constraints and slack variables such as \mathbf{W}_1 , \mathbf{W}_2 , and \mathbf{W}_3 as

$$\begin{aligned} \min_{\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{E}} \quad & \lambda_1 \|\mathbf{W}_1\|_* + \lambda_2 \|\mathbf{W}_2\|_{1,2} \\ & + \lambda_3 \|\mathbf{W}_3\|_{1,1} + \lambda_4 \|\mathbf{E}\|_{1,1} \\ \text{s.t.} \quad & \mathbf{Y} = \mathbf{U}\mathbf{W} + \mathbf{E}, \quad \mathbf{W} = \mathbf{W}_1, \\ & \mathbf{W} = \mathbf{W}_2, \quad \mathbf{W} = \mathbf{W}_3. \end{aligned} \quad (5)$$

The above mentioned problem can be minimized using the conventional inexact augmented Lagrange multiplier (IALM) method [24] that has well-defined convergence properties for nonsmooth optimization which has been used in matrix rank minimization problems. This method is an iterative process that employs the Lagrangian function by adding quadratic penalty terms that assign high cost to the infeasible data and admit closed form updates for each unknown variable. Since the alternating direction method (ADM) [25] is introduced for the low-rank matrix completion and is based on the alternating direction augmented Lagrangian method, we employ ADM for solving minimization of ALM functions.

2.2. Optimization. To solve the objective function given in (5), we formulate the ALM-based ADM model [24, 26] as

$$\begin{aligned} L(\mathbf{W}, \mathbf{E}, \mathbf{M}_{1-4}) &= \lambda_1 \|\mathbf{W}_1\|_* + \lambda_2 \|\mathbf{W}_2\|_{1,2} + \lambda_3 \|\mathbf{W}_3\|_{1,1} + \lambda_4 \|\mathbf{E}\|_{1,1} \\ &+ \langle \mathbf{M}_1, \mathbf{W}_1 - \mathbf{W} \rangle + \langle \mathbf{M}_2, \mathbf{W}_2 - \mathbf{W} \rangle + \langle \mathbf{M}_3, \mathbf{W}_3 - \mathbf{W} \rangle \\ &+ \frac{\alpha}{2} \|\mathbf{W}_1 - \mathbf{W}\|_F^2 + \frac{\alpha}{2} \|\mathbf{W}_2 - \mathbf{W}\|_F^2 + \frac{\alpha}{2} \|\mathbf{W}_3 - \mathbf{W}\|_F^2 \\ &+ \langle \mathbf{M}_4, \mathbf{Y} - \mathbf{U}\mathbf{W} - \mathbf{E} \rangle + \frac{\alpha}{2} \|\mathbf{Y} - \mathbf{U}\mathbf{W} - \mathbf{E}\|_F^2, \end{aligned} \quad (6)$$

where \mathbf{M}_1 , \mathbf{M}_2 , \mathbf{M}_3 , and \mathbf{M}_4 are Lagrangian multipliers, $\|\mathbf{C}\|_F = \sqrt{\sum_i \sum_j |c_{ij}|^2}$ is called Frobenius norm of matrix \mathbf{C} , and $\alpha > 0$ is a penalty parameter.

The proposed strategy for given variable update is performed by computing sparse coefficients based on the low-rank representation with updated error and jointly sparse features. Thus, we first update the sparse error \mathbf{E} of the observation matrix \mathbf{Y} by referring to the proof [24] as

$$\begin{aligned} \mathbf{E}^{k+1} &= \arg \min_{\mathbf{E}} \frac{\lambda_4}{\alpha_k} \|\mathbf{E}\|_{1,1} + \frac{1}{2} \left\| \mathbf{E} - \left(\mathbf{U}\mathbf{W}^k - \mathbf{Y} + \frac{1}{\alpha_k} \mathbf{M}_4^k \right) \right\|_F^2 \\ &= \text{sign}(\mathbf{P}) \max \left(0, 1 - \frac{\lambda_4}{\alpha_k} \right) \mathbf{P}, \end{aligned} \quad (7)$$

where $\mathbf{P} = \mathbf{U}\mathbf{W}^k - \mathbf{Y} + (1/\alpha_k)\mathbf{M}_4^k$.

For updating low-rank variable \mathbf{W}_1 , we employ the proof in [27] such that low-rank matrix approximation problem with F -norm data fidelity can be solved by a soft-thresholding operation on the singular values of observation matrix as

$$\begin{aligned} \mathbf{W}_1^{k+1} &= \arg \min_{\mathbf{W}_1} \frac{\lambda_1}{\alpha_k} \|\mathbf{W}_1\|_* + \frac{1}{2} \left\| \mathbf{W}_1 - \left(\mathbf{W}^k + \frac{1}{\alpha_k} \mathbf{M}_1^k \right) \right\|_F^2 \\ &= \mathbf{U}_R \mathcal{S}_{\lambda_1/\alpha_k}(\Sigma_R), \end{aligned} \quad (8)$$

where $\mathcal{S}_{\lambda_1/\alpha_k}(\Sigma_R)_{ii} = \max(0, (\Sigma_R)_{ii} - (\lambda_1/\alpha_k))$ and $\mathbf{R} = \mathbf{W}^k + (1/\alpha_k)\mathbf{M}_1^k$.

We update the \mathbf{W}_2 and \mathbf{W}_3 referring to the proof [24] as

$$\begin{aligned} \mathbf{W}_2^{k+1} &= \arg \min_{\mathbf{W}_2} \frac{\lambda_2}{\alpha_k} \|\mathbf{W}_2\|_{1,2} + \frac{1}{2} \left\| \mathbf{W}_2 - \left(\mathbf{W}^k + \frac{1}{\alpha_k} \mathbf{M}_2^k \right) \right\|_F^2 \\ &= \max \left(0, 1 - \frac{\lambda_2}{\alpha_k \|\mathbf{W}^k + (1/\alpha_k)\mathbf{M}_2^k\|} \right) \left(\mathbf{W}^k + \frac{1}{\alpha_k} \mathbf{M}_2^k \right), \\ \mathbf{W}_3^{k+1} &= \arg \min_{\mathbf{W}_3} \frac{\lambda_3}{\alpha_k} \|\mathbf{W}_3\|_{1,1} + \frac{1}{2} \left\| \mathbf{W}_3 - \left(\mathbf{W}^k + \frac{1}{\alpha_k} \mathbf{M}_3^k \right) \right\|_F^2 \\ &= \text{sign} \left(\mathbf{W}^k + \frac{1}{\alpha_k} \mathbf{M}_3^k \right) \\ &\quad \times \max \left(0, 1 - \frac{\lambda_3}{\alpha_k} \right) \left(\mathbf{W}^k + \frac{1}{\alpha_k} \mathbf{M}_3^k \right). \end{aligned} \quad (9)$$

The ADM can perform minimization with its alternative property; thus we formulate the objective function for updating sparse coefficient \mathbf{W} based on (6) as

$$\begin{aligned} &\langle \mathbf{M}_1, \mathbf{W} - \mathbf{W}_1 \rangle + \langle \mathbf{M}_2, \mathbf{W} - \mathbf{W}_2 \rangle + \langle \mathbf{M}_3, \mathbf{W} - \mathbf{W}_3 \rangle \\ &+ \frac{\alpha}{2} \|\mathbf{W} - \mathbf{W}_1\|_F^2 + \frac{\alpha}{2} \|\mathbf{W} - \mathbf{W}_2\|_F^2 + \frac{\alpha}{2} \|\mathbf{W} - \mathbf{W}_3\|_F^2 \\ &+ \langle \mathbf{M}_4, \mathbf{W} - \mathbf{U}(\mathbf{Y} - \mathbf{E}) \rangle + \frac{\alpha}{2} \|\mathbf{W} - \mathbf{U}(\mathbf{Y} - \mathbf{E})\|_F^2. \end{aligned} \quad (10)$$

\mathbf{W} is updated with the already updated sparse error \mathbf{E}^{k+1} , low-rank representation \mathbf{W}_1^{k+1} , $\ell_{1,2}$ norm \mathbf{W}_2^{k+1} , and $\ell_{1,1}$ norm \mathbf{W}_3^{k+1} using the following model:

$$\begin{aligned} \mathbf{W}^{k+1} &= \arg \min_{\mathbf{W}} F(\mathbf{W}), \\ \text{s.t.} \quad F(\mathbf{W}) &= \frac{1}{2} \left\| \mathbf{W} - \left(\mathbf{W}_1^{k+1} - \frac{1}{\alpha_k} \mathbf{M}_1^k \right) \right\|_F^2 \\ &+ \frac{1}{2} \left\| \mathbf{W} - \left(\mathbf{W}_2^{k+1} - \frac{1}{\alpha_k} \mathbf{M}_2^k \right) \right\|_F^2 \\ &+ \frac{1}{2} \left\| \mathbf{W} - \left(\mathbf{W}_3^{k+1} - \frac{1}{\alpha_k} \mathbf{M}_3^k \right) \right\|_F^2 \\ &+ \frac{1}{2} \left\| \mathbf{W} - \left(\mathbf{U}^\top (\mathbf{Y} - \mathbf{E}^{k+1}) + \frac{1}{\alpha_k} \mathbf{M}_4^k \right) \right\|_F^2. \end{aligned} \quad (11)$$

We obtain the closed form update framework as

$$\mathbf{W}^{k+1} = \mathcal{F}_{\tau/\delta} \left(\mathbf{W}^k - \frac{1}{2\delta} \nabla F(\mathbf{W}^k) \right), \quad (12)$$

where $\nabla F(\mathbf{W}^k)$ is the partial derivative of $F(\mathbf{W})$ with respect to \mathbf{W}^k and $\mathcal{F}_{\tau/\delta}(x)$ is a component-wise soft-thresholding operator defined by [28] such that if $|x| \leq \tau/\delta$, $\mathcal{F}_{\tau/\delta}(x) = 0$; otherwise, $\mathcal{F}_{\tau/\delta}(x) = x - \text{sign}(x)\tau/\delta$. We set $\tau = 0.25$ and $\delta = 1$.

Then, we update Lagrange multipliers as

$$\begin{aligned} \mathbf{M}_1^{k+1} &= \mathbf{M}_1^k + \alpha_k \left(\mathbf{W}_1^{k+1} - \mathcal{M} \left(\mathbf{W}_1^{k+1} + \frac{\mathbf{M}_1^k}{\alpha_k} \right) \right), \\ \mathbf{M}_2^{k+1} &= \mathbf{M}_2^k + \alpha_k \left(\mathbf{W}_2^{k+1} - \mathcal{M} \left(\mathbf{W}_2^{k+1} + \frac{\mathbf{M}_2^k}{\alpha_k} \right) \right), \\ \mathbf{M}_3^{k+1} &= \mathbf{M}_3^k + \alpha_k \left(\mathbf{W}_3^{k+1} - \mathcal{M} \left(\mathbf{W}_3^{k+1} + \frac{\mathbf{M}_3^k}{\alpha_k} \right) \right), \\ \mathbf{M}_4^{k+1} &= \mathbf{M}_4^k + \alpha_k \left(\mathbf{E}^{k+1} - \mathcal{M} \left(\mathbf{E}^{k+1} + \frac{\mathbf{M}_4^k}{\alpha_k} \right) \right), \end{aligned} \quad (13)$$

where $\mathcal{M}(\mathbf{A})_{ij} = \max(a_{ij}, 0)$.

2.3. Convergence. The convergence of the proposed object tracking algorithm can be guaranteed using the Karush-Kuhn-Tucker (KKT) point approach by referring to linearized ADM with adaptive penalty (LADMMap) approach [26]. The LADMMap [26] aims to solve the following type of convex problems:

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{y}} \quad & f(\mathbf{x}) + g(\mathbf{y}), \\ \text{s.t.} \quad & \mathcal{A}(\mathbf{x}) + \mathcal{B}(\mathbf{y}) = \mathbf{c}, \end{aligned} \quad (14)$$

where \mathbf{x} , \mathbf{y} , and \mathbf{c} could be either vectors or matrices, f and g are convex functions, and \mathcal{A} and \mathcal{B} are linear mappings.

In many problems, f and g are vector or matrix norms, and when \mathcal{A} and \mathcal{B} are identity mappings, the augmented Lagrangian functions of (14) for minimizing \mathbf{x} and \mathbf{y} have closed form solutions. However, when \mathcal{A} and \mathcal{B} are not identity mappings, they may not be easy to solve. Therefore, Lin et al. [26] proposed linearizing the quadratic penalty term in the augmented Lagrangian function and adding a proximal term for updating \mathbf{x} and \mathbf{y} , resulting in the following updating scheme:

$$\begin{aligned} \mathbf{x}_{k+1} &= \arg \min_{\mathbf{x}} f(\mathbf{x}) + \frac{\mu_k \eta_A}{2} \\ &\quad \times \left\| \mathbf{x} - \mathbf{x}_k + \frac{\mathcal{A}^* (\gamma_k + \mu_k (\mathcal{A}(\mathbf{x}_k) + \mathcal{B}(\mathbf{y}_k) - \mathbf{c}))}{(\mu_k \eta_A)} \right\|^2, \\ \mathbf{y}_{k+1} &= \arg \min_{\mathbf{y}} g(\mathbf{y}) + \frac{\mu_k \eta_B}{2} \\ &\quad \times \left\| \mathbf{y} - \mathbf{y}_k + \frac{\mathcal{B}^* (\gamma_k + \mu_k (\mathcal{A}(\mathbf{x}_{k+1}) + \mathcal{B}(\mathbf{y}_k) - \mathbf{c}))}{(\mu_k \eta_B)} \right\|^2, \\ \gamma_{k+1} &= \gamma_k + \mu_k [\mathcal{A}(\mathbf{x}_{k+1}) + \mathcal{B}(\mathbf{y}_{k+1}) - \mathbf{c}], \end{aligned} \quad (15)$$

where γ is the Lagrange multiplier, μ is the penalty parameter, and $\eta_A > 0$ and $\eta_B > 0$ are some parameters for each norm. Lin et al. [26] proposed a strategy to adaptively update the penalty parameter μ and proved that when μ is nondecreasing and is upper bound and $\eta_A > \|\mathcal{A}\|^2$ and $\eta_B > \|\mathcal{B}\|^2$, then $(\mathbf{x}_k, \mathbf{y}_k)$ converges to an optimal solution to (14), where $\|\mathcal{A}\|$ and $\|\mathcal{B}\|$ are the operator norm of \mathcal{A} and \mathcal{B} , respectively.

For updating the penalty parameter α in (6) for the proposed problem, we apply the concept of LADMMap, where \mathbf{x} , \mathbf{y} , \mathbf{c} , and γ are \mathbf{W} , \mathbf{E} , \mathbf{Y} , and \mathbf{M}_{1-4} , respectively, in (6), with some algebra:

$$\alpha_{k+1} = \min(\alpha_{\max}, \rho \alpha_k), \quad (16)$$

where α_{\max} is an upper bound of $\{\alpha_k\}$. The value ρ is decided as

$$\rho = \begin{cases} \rho_0, & \text{if } \alpha_k \cdot \frac{\max(\|\mathbf{E}^{k+1} - \mathbf{E}^k\|_F, \|\mathbf{W}^{k+1} - \mathbf{W}^k\|_F)}{\|\mathbf{Y}\|_F} < \varepsilon_1 \\ 1, & \text{otherwise,} \end{cases} \quad (17)$$

where $\rho_0 \geq 1$ is a constant, and we used $\rho_0 = 1.9$ in this work.

The iterations equations (7)–(13) and (16)–(17) stop when the following two conditions are satisfied:

$$\frac{\|\mathbf{W}^{k+1} + \mathbf{E}^{k+1} - \mathbf{Y}\|_F}{\|\mathbf{Y}\|_F} < \varepsilon_2, \quad (18)$$

$$\alpha_k \cdot \frac{\max(\|\mathbf{E}^{k+1} - \mathbf{E}^k\|_F, \|\mathbf{W}^{k+1} - \mathbf{W}^k\|_F)}{\|\mathbf{Y}\|_F} < \varepsilon_1, \quad (19)$$

where $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ are small thresholds.

The above two stopping criteria are based on the KKT conditions of problem (6). Algorithm 1 summarizes the overall algorithm for the proposed low-rank representation-based multitask feature learning with joint sparsity.

2.4. Object Tracking. Given a set of observed images at the t th frame $\mathcal{Y}_t = \{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_t\}$, the goal of object tracking is to recursively estimate the state \mathbf{x}_t , maximizing the following state distribution:

$$p(\mathbf{x}_t | \mathcal{Y}_t) \propto p(\mathbf{Y}_t | \mathbf{x}_t) \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x} | \mathcal{Y}_{t-1}) d\mathbf{x}_{t-1}, \quad (20)$$

where $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ denotes the motion model between two temporally consecutive states and $p(\mathbf{Y}_t | \mathbf{x}_t)$ indicates the likelihood function.

In the motion model, we regard the state variable \mathbf{x}_t as six affine parameters [4] such as horizontal and vertical translations, rotation, scale, aspect ratio, and skew such as $\mathbf{x}_t = [x_t, y_t, \theta_t, s_t, \beta_t, \phi_t]$. The motion model is assumed to have random Gaussian distribution such as $p(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \mathbf{x}_{t-1}, \Phi)$, where Φ is a diagonal covariance matrix.

The solution of the proposed algorithm in Algorithm 1 can be obtained by minimizing the following functional as

$$\min_{\mathbf{W}^{(i)}, \mathbf{E}^{(i)}} \frac{1}{2} \|\mathbf{Y}^{(i)} - \mathbf{U}\mathbf{W}^{(i)} - \mathbf{E}^{(i)}\|_F^2 + \beta \|\mathbf{E}^{(i)}\|_{1,1}, \quad (21)$$

(1) Input: A data set of n data points $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]$, and parameters $\lambda_1, \lambda_2, \lambda_3, \lambda_4$.
(2) Output: Optimal coefficient \mathbf{W}_{opt} , and errors \mathbf{E}_{opt}
(3) Initialization: Set $\mathbf{W}^0 = \mathbf{0}, \mathbf{E}^0 = \mathbf{0}, \mathbf{M}_1^0 = \mathbf{0}, \mathbf{M}_2^0 = \mathbf{0}, \mathbf{M}_3^0 = \mathbf{0}, \mathbf{M}_4^0 = \mathbf{0}, \mathbf{W}_1^0 = \mathbf{0}, \mathbf{W}_2^0 = \mathbf{0}, \mathbf{W}_3^0 = \mathbf{0}, \varepsilon_1 < \varepsilon_2$, and $0 < \alpha_0 \ll \alpha_{\text{max}}$
(4) **Iterate until (18) and (19) is not satisfied:**
Fix other variables and update \mathbf{E} using (7)
Fix other variables and update \mathbf{W}_1 using (8)
Fix other variables and update \mathbf{W}_2 and \mathbf{W}_3 using (9) and (10)
Fix other variables and update \mathbf{W} using (12)
Update $\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_4$ and parameters using (13)
Update α using (16) and (17)
(5) **End**

ALGORITHM 1: The algorithm for low-rank representation-based multitask feature learning.

where i indicates the i th sample data of the state \mathbf{x} and a reconstruction error for the nonoccluded region of the object is given as $\mathbf{Y}^{(i)} - \mathbf{U}\mathbf{W}^{(i)} - \mathbf{E}^{(i)}$.

The likelihood function can be formulated by the reconstruction error given in [29]

$$p(\mathbf{Y}^{(i)} | \mathbf{x}^{(i)}) = \exp\left(-\|\mathbf{Y}^{(i)} - \mathbf{U}\mathbf{W}^{(i)}\|_F^2\right). \quad (22)$$

This likelihood function, however, cannot deal with severe appearance deformations. For this reason, the likelihood function is modified by an additionally labeled penalty constraint about the similarity in neighboring error matrices for the deformed region of the object such as $|(1 - \mathbf{A}^{(i+1)}) - (1 - \mathbf{A}^{(i)})|$, where $\mathbf{A}^{(i)}$ and $\mathbf{A}^{(i+1)}$, respectively, represent matrices with nonzero elements of $\mathbf{E}^{(i)}$ and $\mathbf{E}^{(i+1)}$. Thus, the proposed likelihood function is given as

$$\begin{aligned} p(\mathbf{Y}^{(i)} | \mathbf{x}^{(i)}) &= \exp\left[-\left\|\mathbf{A}^{(i)} \odot (\mathbf{Y}^{(i)} - \mathbf{U}\mathbf{W}^{(i)})\right\|_F^2\right. \\ &\quad \left. + \beta \sum \left| (1 - \mathbf{A}^{(i+1)}) - (1 - \mathbf{A}^{(i)}) \right| \right], \end{aligned} \quad (23)$$

where \odot is the element-wise multiplication of matrices.

3. Experimental Results

The proposed object tracking algorithm is implemented and tested using ten challenging sequences. Major challenging issues in the test sequences include scale variation, shape deformation, fast motion, out-of-plane rotation, background clutter, object rotation, occlusion, illumination variation, out of view, motion blur, and low resolution. The proposed method is compared with a number of state-of-the-art tracking algorithms such as SMTT [15], APGLI [14], SPCA [21], ASL [2], ILRF [18], and RMTT [16].

The proposed object tracking algorithm is implemented in MATLAB and processes 1.5 frames per second on a Pentium 2.7 GHz dual core PC without any hardware accelerator such as GPU. For each test sequence, the initial location of the object is manually selected in the first frame. Each image

sample from the target and background is normalized to a 32×32 patch.

3.1. Regularization Parameters $\lambda_1, \lambda_2, \lambda_3$, and λ_4 . Regularization parameters in (4) are empirically chosen using the center pixel error criterion in [30].

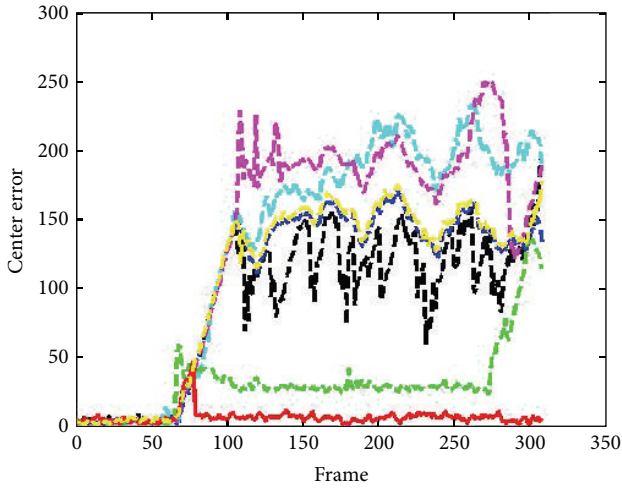
We fix $\lambda_4 = 1.0$ and swap others since λ_4 represents the sparse error matrix \mathbf{E} while λ_1, λ_2 , and λ_3 are related to weight variables $\mathbf{W}_1, \mathbf{W}_2$, and \mathbf{W}_3 , respectively. Parameters λ_1, λ_2 , and λ_3 are parameterized by a discrete set with 10 elements as $\{1e^{-4}, 1e^{-3}, 0.01, 0.1, 0.5, 1.0, 3.0, 5.0, 10.0, 15.0\}$.

We test different combinations of λ_1, λ_2 , and λ_3 about three different images sequences having challenging issues such as occlusion, deformation, and scale variation with 1500 frames and compute the central pixel errors from all frames. For each λ_1 in the discrete set with different λ_2 and λ_3 , we produce 10×10 average center pixel errors. When tracking performance is acceptable for giving the fixed λ_1 , we average all the 10×10 center pixel errors. With this attempt, we set the regularization parameters as $\lambda_1 = 5, \lambda_2 = 0.5, \lambda_3 = 0.1$, and $\lambda_4 = 1.0$.

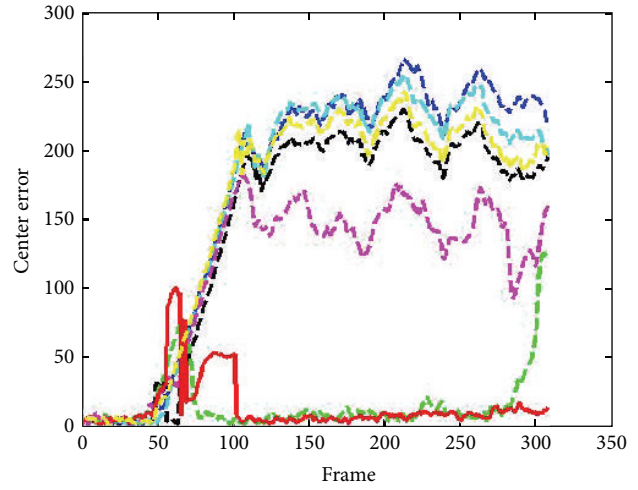
3.2. Quantitative Evaluation. For quantitative performance comparison, center pixel error evaluation method is used and overlap ratio criterion is computed. The center pixel error represents the distance between the predicted and the ground truth center pixels. Figure 1 shows the result of center pixel errors of seven different object tracking algorithms for ten test sequences.

The overlap ratio criterion represents the ratio between the number of frames for a specific object to be completely tracked and the total number of frames in the image sequence. In order to decide whether the object is successfully tracked, we employ the overlapped score defined in [30]. Figure 2 shows overlap ratios between the ground truth region and the tracking region.

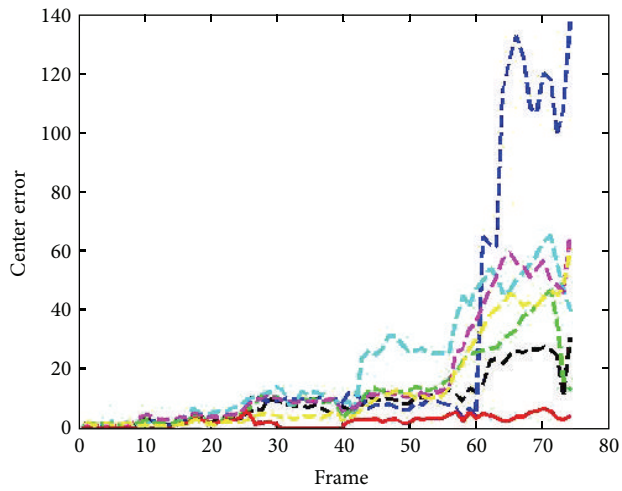
3.3. Qualitative Evaluation. Jogging 1 and Jogging 2 sequences include a variety of critical conditions, such as disappearing object, out-of-plane rotation, and deformation. Object deformation and out-of-plane rotation make all the



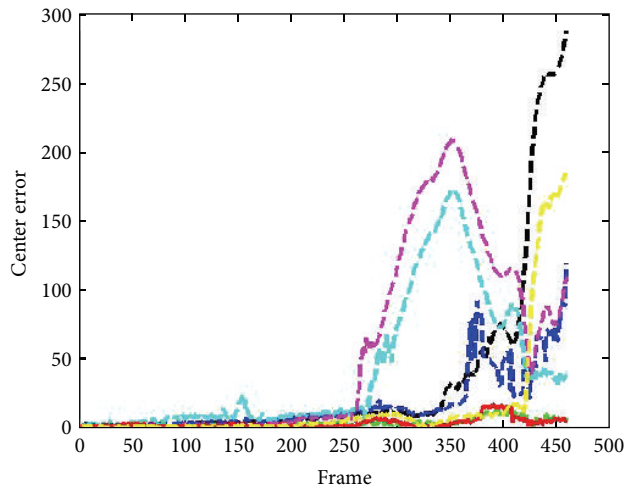
(a) Jogging 1



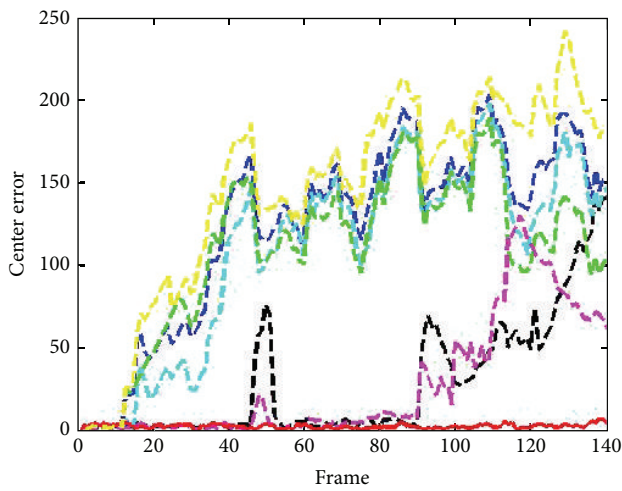
(b) Jogging 2



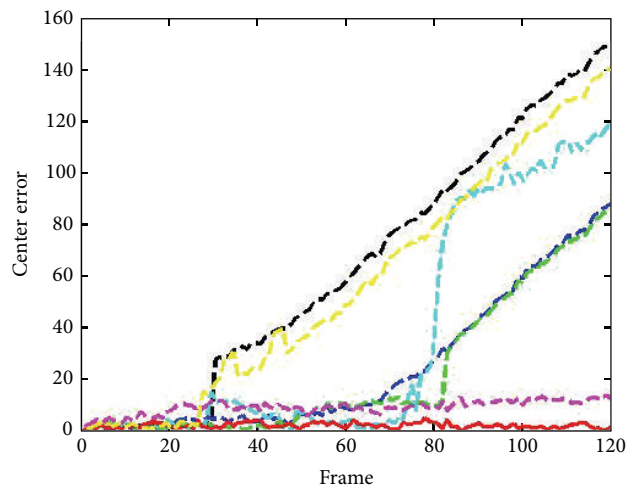
(c) Football 1



(d) Freeman 3



(e) Couple



(f) Crossing

--- MTT
--- APGL1
--- SPCA
--- ASL
--- ILRFT
--- RMTT
--- Proposed

--- MTT
--- APGL1
--- SPCA
--- ASL
--- ILRFT
--- RMTT
--- Proposed

FIGURE 1: Continued.

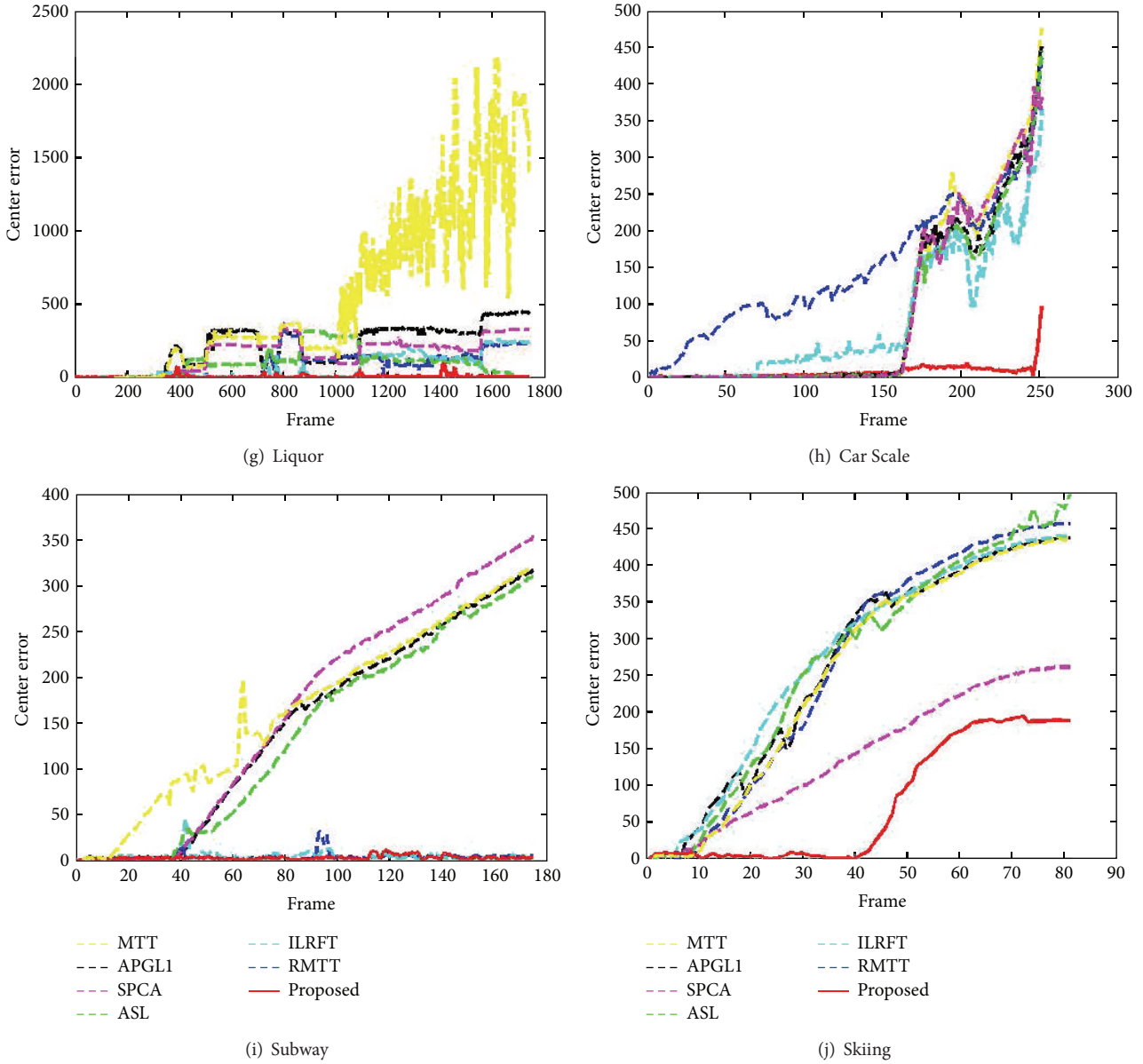


FIGURE 1: Center pixel error. This figure shows center pixel error for ten test sequences. The proposed method is compared with six state-of-the-art methods.

existing methods except ASL [2] drift away object. On the other hand, the proposed tracking method can successfully track the object since it can represent the object appearance more completely using a large scale training data with efficient multitask feature learning for object representation.

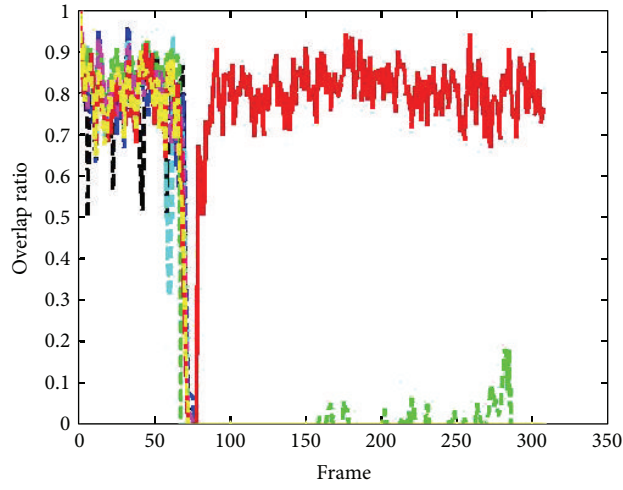
The moving object in the Football 1 sequence undergoes in- and out-of-plane rotation and background clutter. Experimental results demonstrate that the proposed method achieves the best tracking performance in this sequence. Other methods cannot avoid drift at some instances in the neighborhood of the 60th frame.

There is a blurry scale variation in Freeman 3 and Car Scale sequences. For this reason, it is difficult to predict the location of the moving object. Furthermore, it includes

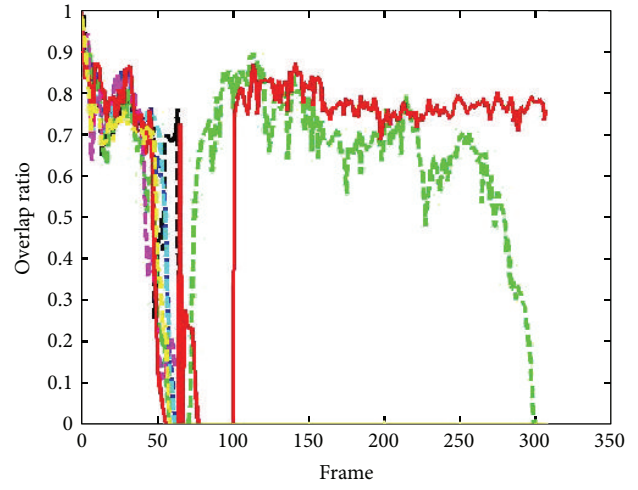
drastic appearance change caused by motion blur. The proposed method performs well since it can adapt to the scale and appearance change of the object and overcome the influence of motion blur by using joint sparsity-based object appearance representation.

In addition to scale variation and in- and out-of-plane object rotations, Couple and Crossing sequences include a critical factor such as fast motion. The fast motion can be involved by both camera and object. For tracking an object in Crossing sequence, the proposed method and SPCA [21] methods perform very well whereas other methods lost their object in neighborhood of the 30th and 80th frames.

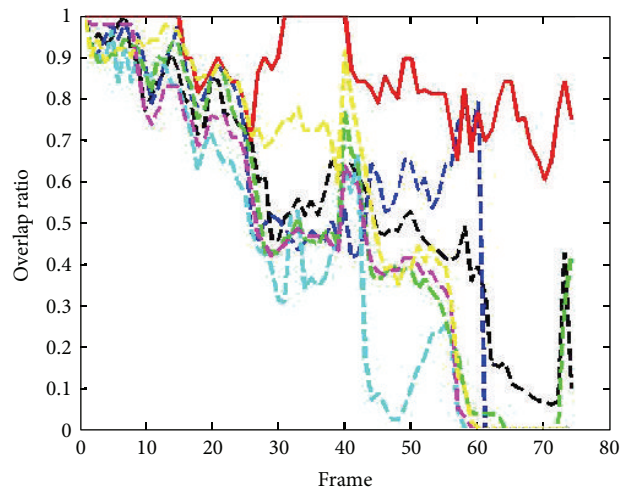
The Liquor sequence contains severe motion blur, scale and illumination variation, and fast motion. Although full



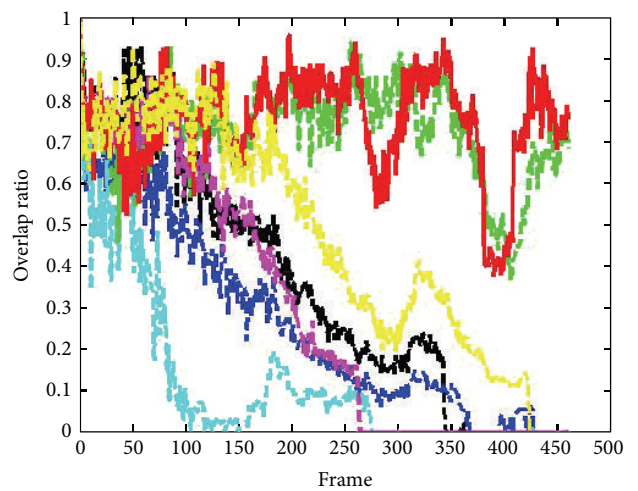
(a) Jogging 1



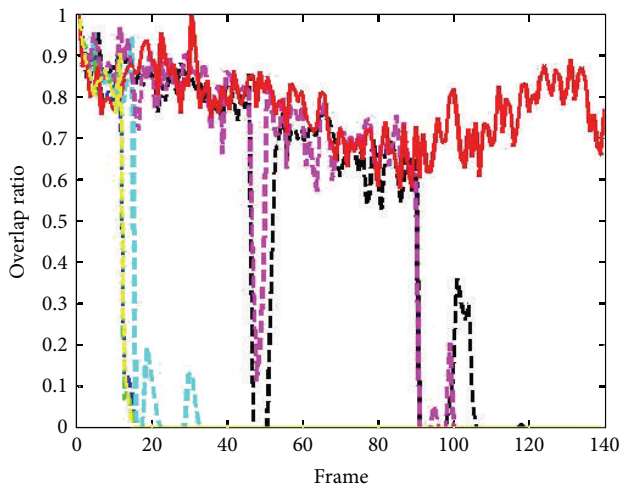
(b) Jogging 2



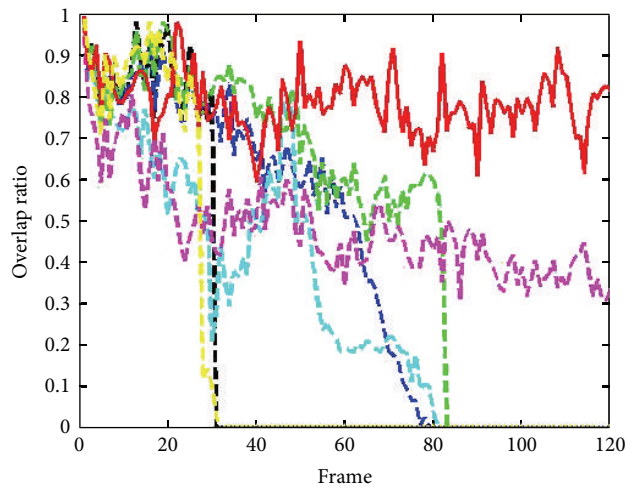
(c) Football 1



(d) Freeman 3



(e) Couple



(f) Crossing

--- MTT
--- APGL1
--- SPCA
--- ASL
--- ILRFT
--- RMTT
--- Proposed

--- MTT
--- APGL1
--- SPCA
--- ASL
--- ILRFT
--- RMTT
--- Proposed

FIGURE 2: Continued.

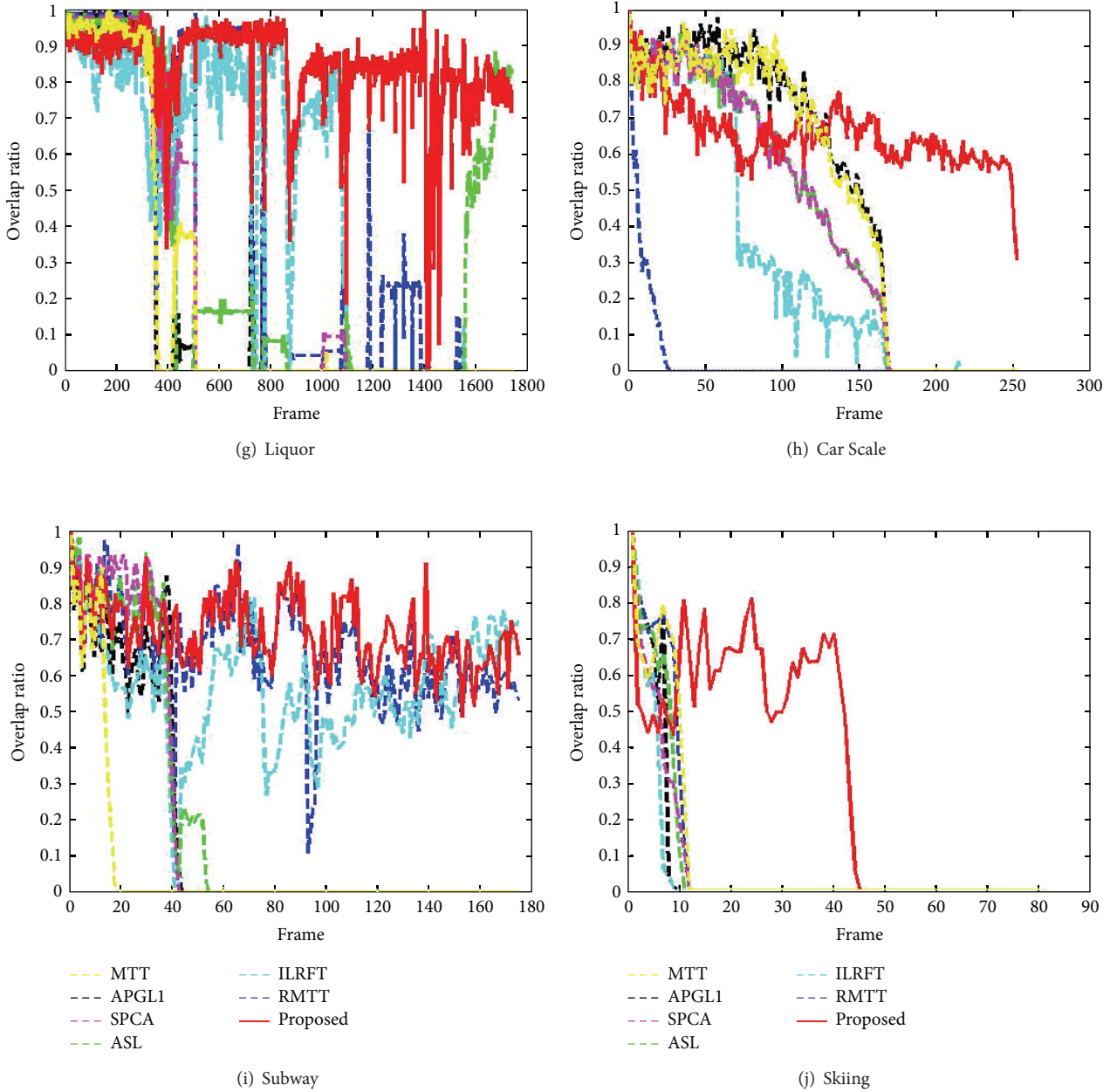


FIGURE 2: Overlap ratio evaluation. This figure shows overlap ratios for ten test sequences. The proposed method is compared with six state-of-the-art methods.

occlusion occurs by a bottle with the similar color, the proposed tracking method successfully tracks the object without drifting.

In the Subway sequence, object’s information is lost by discriminant factor such as background clutter and the total occlusion. The proposed method can successfully track the moving object since it preserves the sparsity of object appearance with optimal sparse coding using low-rank feature representation with considering the joint sparsity. Other trackers except ILRFT [18] and RMTT [16] drift in the neighborhood of the 20th and 40th frames.

The Skiing sequence contains severely deformed object. The proposed method successfully performs tracking due to its novel object likelihood function defined in (21). It employs more refined likelihood function formulated by the weighted reconstruction error for rigid object regions and by the labeling penalty constraint about the similarity in residual matrices for deformed object regions as well. Figure 3 shows the overall tracking results in each test sequence. The overlapped ratio between the ground truth and predicted tracking region using proposed method turns to 0 at around frame 40, while the center pixel error between the ground

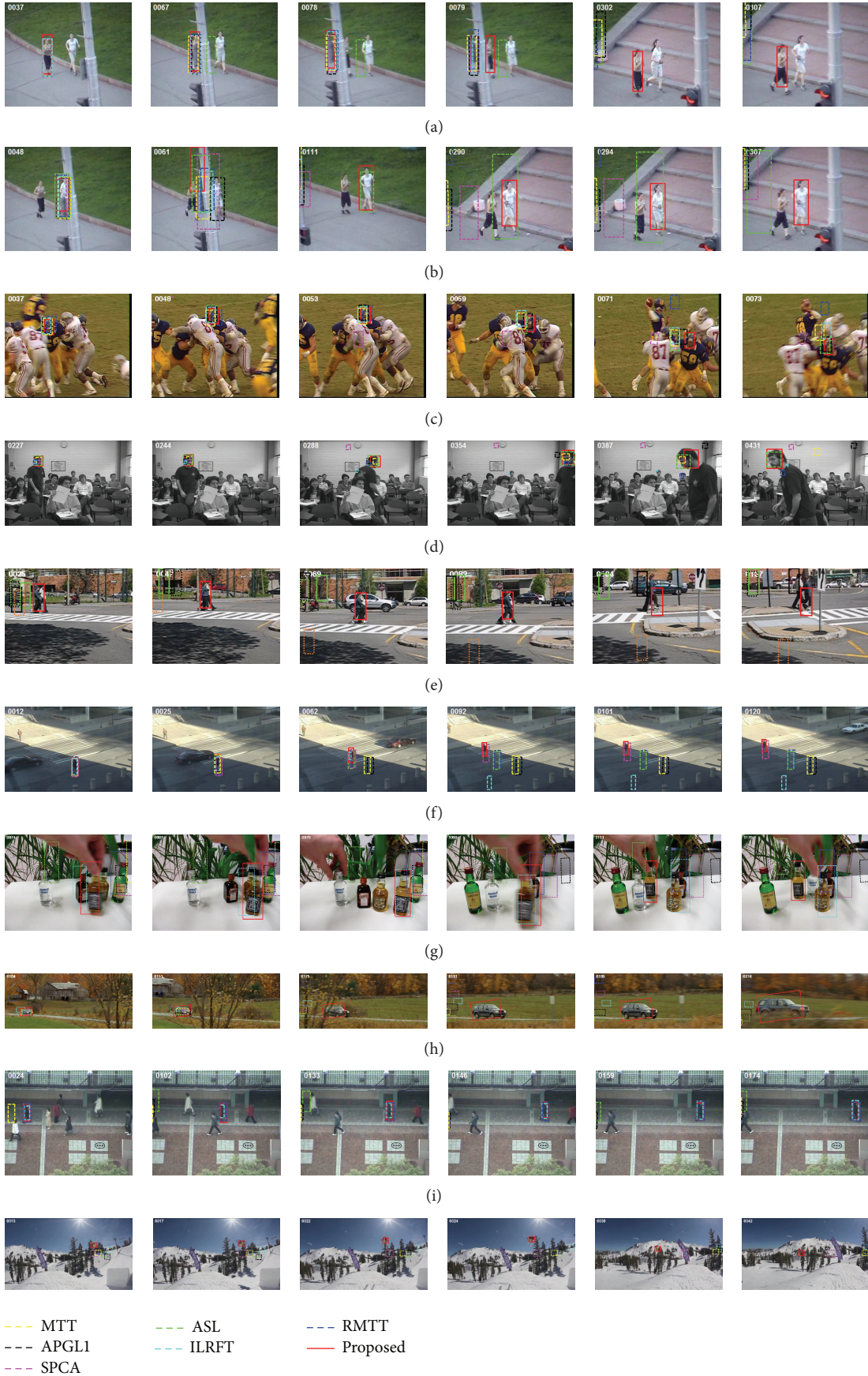


FIGURE 3: Object tracking results using seven different algorithms on ten test sequences. They are Jogging 1, Jogging 2, Football 1, Freeman 3, Couple, Crossing, Liquor, Car Scale, Subway, and Skiing, respectively, from the top row to the bottom row.

truth and predicted tracking region using proposed method still increases. From this fact, we conclude that the proposed method loses tracking in around frame 40, while other state-of-the-art tracking methods totally lose their tracking in around frame 10.

4. Conclusion

In this paper, we present an effective, robust object tracking method using a multitask feature learning-based low-rank representation with joint sparsity. In order to overcome the limitation of existing sparse representation-based object tracking methods, we employ the novel optimization process of low-rank representation of objects by using a recently proposed model minimization method. The efficient subspace learning-based sparse coding and simultaneous update method of both optimal sparse codes and error matrix can be appropriately updated in the process of tracking in case of severe appearance variation. Experimental results demonstrate that the proposed method can successfully track objects in various video sequences with critical issues such as occlusion, deformation, plane rotations, background clutter, motion blur, scale and illumination variations, and fast motion.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the MSIP, Korea, under the ITRC Support Program (NIPA-2014-CAU) supervised by NIPA, in part by the Technology Innovation Program (Development of Smart Video/Audio Surveillance SoC and Core Component for Onsite Decision Security System) under Grant 10047788, and by the Ministry of Science, ICT and Future Planning as Software Grand Challenge Project (Grant no. 14-824-09-002).

References

- [1] S. Wang, H. Lu, F. Yang, and M.-H. Yang, "Superpixel tracking," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 1323–1330, IEEE, November 2011.
- [2] X. Jia, H. Lu, and M.-H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 1822–1829, Providence, RI, USA, June 2012.
- [3] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–577, 2003.
- [4] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *International Journal of Computer Vision*, vol. 77, no. 1–3, pp. 125–141, 2008.
- [5] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, vol. 1, pp. 798–805, June 2006.
- [6] W. Hu, X. Li, X. Zhang, X. Shi, S. Maybank, and Z. Zhang, "Incremental tensor subspace learning and its applications to foreground segmentation and tracking," *International Journal of Computer Vision*, vol. 91, no. 3, pp. 303–327, 2011.
- [7] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," in *Computer Vision—ECCV*, pp. 234–247, Springer, 2008.
- [8] S. Avidan, "Support Vector Tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1064–1072, 2004.
- [9] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2259–2272, 2011.
- [10] X. Mei, H. Ling, Y. Wu, E. Blasch, and L. Bai, "Minimum error bounded efficient ℓ_1 tracker with occlusion detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 1257–1264, Providence, RI, USA, June 2011.
- [11] B. Liu, L. Yang, J. Huang, P. Meer, L. Gong, and C. Kulikowski, "Robust and fast collaborative tracking with two stage sparse optimization," in *Proceedings of the 11th European Conference on Computer Vision (ECCV '10)*, pp. 624–637, Springer, 2010.
- [12] B. Liu, J. Huang, C. Kulikowski, and L. Yang, "Robust visual tracking using local sparse appearance model and k-selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2968–2981, 2013.
- [13] H. Kim and J. Paik, "Object tracking using compressive local appearance model with ℓ_1 -regularisation," *Electronics Letters*, vol. 50, no. 6, pp. 444–446, 2014.
- [14] C. Bao, Y. Wu, H. Ling, and H. Ji, "Real time robust ℓ_1 tracker using accelerated proximal gradient approach," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 1830–1837, IEEE, June 2012.
- [15] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via structured multi-task sparse learning," *International Journal of Computer Vision*, vol. 101, no. 2, pp. 367–383, 2013.
- [16] Y. Bai and M. Tang, "Object tracking via robust multitask sparse representation," *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 909–913, 2014.
- [17] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Low-rank sparse learning for robust visual tracking," in *Computer Vision—ECCV 2012*, pp. 470–484, Springer, 2012.
- [18] C. Zhang, R. Liu, T. Qiu, and Z. Su, "Robust visual tracking via incremental low-rank features learning," *Neurocomputing*, vol. 131, pp. 237–247, 2014.
- [19] T. Zhang, S. Liu, N. Ahuja, M.-H. Yang, and B. Ghanem, "Robust visual tracking via consistent low-rank sparse learning," *International Journal of Computer Vision*, pp. 1–20, 2014.
- [20] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, article 11, 2011.
- [21] D. Wang, H. Lu, and M.-H. Yang, "Online object tracking with sparse prototypes," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 314–325, 2013.
- [22] A. Argyriou, T. Evgeniou, and M. Pontil, "Multi-task feature learning," in *Advances in Neural Information Processing Systems*, vol. 19, pp. 41–48, MIT Press, 2007.

- [23] A. Argyriou, T. Evgeniou, and M. Pontil, "Convex multi-task feature learning," *Machine Learning*, vol. 73, no. 3, pp. 243–272, 2008.
- [24] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," In press, <http://arxiv.org/abs/1009.5055>.
- [25] Y. Xu, W. Yin, Z. Wen, and Y. Zhang, "An alternating direction algorithm for matrix completion with nonnegative factors," *Frontiers of Mathematics in China*, vol. 7, no. 2, pp. 365–384, 2012.
- [26] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Advances in Neural Information Processing Systems*, pp. 612–620, 2011.
- [27] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [28] K. Bredies and D. A. Lorenz, "Linear convergence of iterative soft-thresholding," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 813–837, 2008.
- [29] T. Wang, I. Y. H. Gu, and P. Shi, "Object tracking using incremental 2D-PCA learning and ML estimation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, vol. 1, pp. I-933–I-936, April 2007.
- [30] M. Everingham, L. van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.