Sigma Set Based Implicit Online Learning for Object Tracking

Xiaopeng Hong, Hong Chang, *Member, IEEE*, Shiguang Shan, *Member, IEEE*, Bineng Zhong, Xilin Chen, *Senior Member, IEEE*, and Wen Gao, *Fellow, IEEE*

***Abstract*—This letter presents a novel object tracking approach within the Bayesian inference framework through implicit online learning. In our approach, the target is represented by multiple patches, each of which is encoded by a powerful and efficient region descriptor called Sigma set. To model each target patch, we propose to utilize the online one-class support vector machine algorithm, named Implicit online Learning with Kernels Model (ILKM). ILKM is simple, efficient, and capable of learning a robust online target predictor in the presence of appearance changes. Responses of ILKMs related to multiple target patches are fused by an arbitrator with an inference of possible partial occlusions, to make the decision and trigger the model update. Experimental results demonstrate that the proposed tracking approach is effective and efficient in ever-changing and cluttered scenes.**

***Index Terms*—Object tracking, implicit online learning with kernels, particle filter, Sigma set.**

I. Introduction

OBJECT tracking is to estimate the status of moving objects in image sequences. Tracking plays an important role in automatic visual surveillance, motion analysis, human-computer interaction and traffic monitoring, etc. Although tremendous success has been achieved in the tracking literature, robust, accurate, and real-time object tracking in unconstrained environment remains a very challenging problem, due to the difficulties coming from occlusion, appearance changes of moving objects, cluttered scenes, etc.

To conquer these difficulties, many tracking methods have been proposed in the last decades [22]. Recently, most of the top-performing approaches rely on the online models through online learning approaches [1], [2], [5]–[8], [10], [11], [13]–[17], [19]–[21] to adapt to the changes in the target appearance or contour, etc. Plenty of sophisticated features or descriptors, such as contour or silhouettes [9], [15], colors [1], [5], [13], [14], [21], Haar-like features [2], [6], [11], texture [15], [17], gradients [1], [7], [13], [14], and joint spatial-color space [19] are adopted to form discriminative target representation. Among them, the descriptors based on second order statistics achieve promising results for tracking [8], [10], [16], [20]. Porikli *et al.* propose to adopt the covariance matrices (COVs) as the object appearance models [16], which outperform the color histograms. COVs capture the correlations among extracted features inside image regions (patches) covered by objects in low dimensionality and are robust against the illumination variations and noises, etc. Despite such advantages, because covariance matrices do not lie in Euclidean space, computationally demanding operations based on Riemannian manifold are required to measure the distance between two COVs and calculate the mean of COVs accurately.

Actually, real-world object tracking approaches are expected to possess not only good performance but also high efficiency. To improve the efficiency of covariance tracking, Li *et al.* [10] and Wu *et al.* [20] utilize the log-Euclidean metric that is relatively more efficient than the affine-invariant metric in [16] but requires the operations on Riemannian manifold as well. Recently, we design the Sigma set region descriptor and apply it to object tracking [8]. Sigma set implicitly encodes the feature correlations in the form of a point set and can be efficiently evaluated. It is demonstrated that Sigma set tracking achieves higher efficiency than the covariance tracking in the calculation of the distance between two descriptors and the mean of descriptors, while preserving similar accuracy [8].

This letter further extends the work in [8] to adapt it to the appearance changes of moving objects promptly. We propose to utilize the Implicit online Learning with Kernels Model (ILKM)[1] [4], which is a kernel-based online one-class Support Vector Machine (1-SVM), to model each target patch separately. ILKM learns a robust predictor by assigning large weights to the samples around the boundary of the class and last observed, i.e., the samples relevant to the online discriminative task. It is simple to implement and efficient for both the temporal and spatial requirement. Moreover, we empirically show that using it as the patch model based on Sigma set in the commonly used Bayesian inference framework achieves promising results in object tracking.

Fig. 1 illustrates the proposed appearance model. In Fig. 1(a), an object is represented by multiple patches, each described by a Sigma set descriptor. In Fig. 1(b), each patch of the object is modeled by an ILKM with its Sigma set and an arbitrator is used to make the final decision on the responses of ILKMs.

Manuscript received April 18, 2010; revised June 17, 2010; accepted June 21, 2010. Date of publication July 12, 2010; date of current version July 22, 2010. This work was performed at Institute of Computing Technology (ICT) and was supported in part by NSFC under Contracts U0835005, 60832004, 60872124; and by the Grand Program of International S&T Cooperation of Zhejiang Province S&T Department under Contract 2008C14063. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jen-Tzung Chien.

X. Hong and B. Zhong are with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China (e-mail: xphong@jdl.ac.cn; bnzhong@jdl.ac.cn).

H. Chang, S. Shan, and X. Chen are with Key Laboratory of Intelligent Information Processing, ICT, Chinese Academy of Sciences, Beijing, China (e-mail: hchang@jdl.ac.cn; xlchen@jdl.ac.cn).

W. Gao is with Institute of Digital Media, Peking University, Beijing, China (e-mail: wgao@jdl.ac.cn).

Color versions of one of more figures in this paper are available at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2010.2057507

[1]"*Implicit*" refers to the manner of solving the risk function of ILKM.

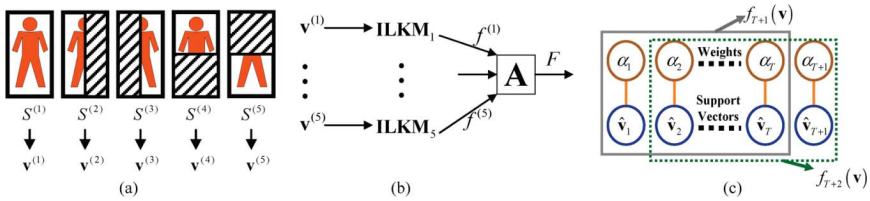

Fig. 1. Illustration of the object model. (a) Five patches of an object, where Sigma set $S^{(n)}$ of the nth patch is represented as a vector $\mathbf{v}^{(n)}$; (b) an arbitrator (denoted as "A") makes the decision based on the responses of the five ILKMs corresponding to the five patches; and (c) Model update of a "sparse" ILKM with length T. A new observation at time $T+1$ is used for update whereas the last observed one at time 1 is removed. See text for more details.

In Fig. 1(c), a reliable observation at current time is utilized to update the ILKM while the earliest observed one is removed.

The benefits of the proposed appearance model are fourfold: *firstly*, patch based representation increases the robustness against partial occlusions; *secondly*, Sigma set preserves the discriminative power and robustness of tradition COVs with higher efficiency; *thirdly*, ILKM is capable of learning a robust predictor based on the observed samples according to their confidences efficiently; *finally*, the arbitrator including an analysis of possible partial occlusions further improves the robustness.

II. Sigma Set Based Appearance Model

A. Multi-Patch Object Representation

To handle partial occlusions and increase the discriminative power as well, a *target* is represented by multiple patches [10], [20]. Fig. 1(a) illustrates a five-patch example, where the *whole* region captures the global information of the target, and the *left*, *right*, *top*, and *bottom* halves encode the local information. Each patch is then represented by a Sigma set region descriptor [8]. Sigma set is formed by a set of constructed points (i.e., feature vectors), which owns the same 2nd order statistics as the given patch. Hence the correlations among extracted features inside a patch are encoded implicitly by Sigma sets, rather than represented directly using COVs.

As COVs, Sigma sets are based on the statistics of pixel-wise feature vectors inside an image region. Hence the selection of elementary features is crucial. To capture the color information when available, this letter utilizes the CIE LAB color space [3] (denoted as LAB), rather than the traditional RGB color space [16]. Since LAB is designed to approximate human vision uniformity, using the Euclidean distance in LAB coordinates to measure the color difference is more consistent with the human perception than in the device-dependent RGB space. In particular, seven elementary features are adopted:

$$\mathbf{f}(x,y) = [L, a, b, x, y, |L_x|, |L_y|]^T \tag{1}$$

where the intensity L, two color-opponent dimensions a and b are the LAB coordinates of the pixel at position (x,y), while L_x and L_y denote the 1st order derivatives of L, respectively.

Given the $d=7$ dimensional feature vectors in (1) inside an image region R, the $d\times d$ covariance matrix $\mathbf{C}$ of R is calculated [18]. Then Sigma set S can be constructed from the d columns of $\mathbf{L} = (\mathbf{L}_1, \ldots, \mathbf{L}_d)$, where $\mathbf{L}$ is the lower triangular matrix square root of $\mathbf{C}$, i.e., $\mathbf{L}\mathbf{L}^T = \mathbf{C}$, obtained through Cholesky factorization [8]. We refer readers to the related papers [18], [8] for more details of COVs and Sigma sets construction.

This letter simplifies the definition of Sigma set as follows:

$$S = \{\mathbf{L}_1, \ldots, \mathbf{L}_d\}. \tag{2}$$

It is slightly different from the original form in [8], which consists of additional d elements with minus signs and each element of which is multiplied by a constant [8]. Since Sigma set is derived from COV uniquely, the elegant robustness and discriminative power of the covariance matrix are directly inherited by Sigma set [8].

Using Sigma sets as descriptors are more efficient than using COVs [8], since the distance between two Sigma sets can be calculated through the distances between the points of these two sets, which are defined in the vector space. In this letter, we utilize the point restricted modified Hausdorff distance (i.e., (7) in [8]) as the distance between Sigma sets. Formally, the distance between two Sigma sets S_A and S_B of two patches A and B (when exploiting Manhattan distance to measure the differences between the points of Sigma sets), is computed as:

$$\rho(S_A, S_B) = \|\mathbf{v}_A - \mathbf{v}_B\|_1 \tag{3}$$

where $\mathbf{v}$ is a $(d^2+d)/2 = 28$ dimensional vector concatenating all the lower triangular entries of $\mathbf{L}$ and $\|.\|_1$ denotes the L1-norm of a vector. The computational complexity of (3) is $O(d^2)$, much lower than that of the Riemannian manifold based metrics of COVs [10], [16], where the time-consuming matrix operations such as matrix logarithm (of computational complexity $O(d^3)$) are required. We refer readers to the related paper [8] for more details of the efficiency comparison [8].

With the above discussion, we use $\mathbf{v}$ to denote a Sigma set as well as the patch hereinafter.

B. ILKM Based Appearance Model

Directly using the mean of T previous descriptors as the online appearance model [8], [16] may easily cause the *drift* problem [15] because errors between the true states and candidates would be accumulated inevitably. Hence we utilize the Implicit online Learning with Kernels Models (ILKMs) [4] to adapt to the appearance changes of each object patch promptly, and adopt an arbitrator to make the final decision and trigger the model update (which will be described in Section II-C).

ILKM is a kernel-based online learning algorithm derived from 1-SVM [4]. The learned predictor of ILKM is a weighted combination of a series of kernel functions referring to the previous observations. Large weights are assigned to the samples lying around the boundary and last observed according to the maximum margin principle. ILKM has been successfully applied to several computer vision tasks, such as background subtraction [4].

Specifically, given the descriptor $\mathbf{v}_t^{(n)}$ of the nth patch of a candidate at time $t, n = 1, \ldots, N$, its confidence belonging to the *target* is given by the nth ILKM as follows:

$$f_t^{(n)}\left(\mathbf{v}_t^{(n)}\right) = \sum_{i=1}^{t-1} \alpha_i^{(n)} k\left(\hat{\mathbf{v}}_i^{(n)}, \mathbf{v}_t^{(n)}\right) \tag{4}$$

where $k(.,.)$ is a kernel function. $\hat{\mathbf{v}}_i^{(n)}$ and $\alpha_i^{(n)} \geq 0, i = 1,\ldots,t-1$ are respectively the support vectors obtained from the historical observations and their corresponding weights.

Stump kernel [12] in (5) is utilized as the kernel function in (4), in consideration that 1) it performs similarly to the radial basis function kernel [12]; 2) it is simple and efficient; and 3) it is consistent with the distance between Sigma sets in (3).

$$k(\mathbf{v}_A, \mathbf{v}_B) = C - ||\mathbf{v}_A - \mathbf{v}_B||_1 \tag{5}$$

where $\mathbf{v}_A$ and $\mathbf{v}_B$ are two vectors as mentioned above, and C is a constant (Since C only shifts the function response, it is set to 1 in this letter).

To fuse the N confidences obtained by (4) and make the final decision, an arbitration procedure is utilized, as illustrated in Fig. 1(b). The basic idea is to neglect the patch of the candidate most dissimilar to the target, which has high possibility to suffer a sudden appearance change or occlusion, etc. A similar strategy is applied in object detection successfully [18]. In particular, given the concatenation of the N patch descriptors $\mathbf{z}_t = [\mathbf{v}_t^{(1)}; \ldots; \mathbf{v}_t^{(N)}]$ at time t, the fused confidence is calculated in the following manner:

$$F_t(\mathbf{z}_t) = \sum_{n=1}^{N} f_t^{(n)}\left(\mathbf{v}_t^{(n)}\right) - \min_{m=1,\cdots,N}\left[f_t^{(m)}\left(\mathbf{v}_t^{(m)}\right)\right]. \tag{6}$$

C. Implicit Online Update

The nth ILKM is updated using $\hat{\mathbf{v}}_t^{(n)}$, the nth patch of the newly estimated candidate $\hat{\mathbf{z}}_t$ at time t in the following manner:

$$f_{t+1}^{(n)}(\mathbf{v}^{(n)}) = (1-\tau) f_t^{(n)}\left(\mathbf{v}^{(n)}\right) + \alpha_t^{(n)} k\left(\hat{\mathbf{v}}_t^{(n)}, \mathbf{v}^{(n)}\right) \tag{7}$$

where $\tau \in [0,1)$ is the damping rate to down weight those historical samples and $\alpha_t^{(n)}$ is the coefficient of $\hat{\mathbf{v}}_t^{(n)}$ online learnt by a monotonic decreasing function of the confidence of $\hat{\mathbf{v}}_t^{(n)}$, as formulated as follows:

$$\alpha_t^{(n)} = \max\left(\frac{1-(1-\tau) f_t^{(n)}\left(\hat{\mathbf{v}}_t^{(n)}\right)}{k\left(\hat{\mathbf{v}}_t^{(n)}, \hat{\mathbf{v}}_t^{(n)}\right)}, 0\right). \tag{8}$$

In [4], ILKM is updated directly once a new observation is obtained. However, to avoid assigning large coefficients to outliers, here we only update the model when acquiring a sufficiently "reliable" candidate by setting a threshold tr on the arbitrator response F_t in (6). In other word, when and only when $F_t(\cdot) > \mathrm{tr}$, each of the N ILKMs is updated using (7).

In real time object tracking, storing all the historical observations up to time $t-1$ like (4) is prohibitively expensive. Therefore, we adopt the "sparse ILKM" (SILK) [4] to reduce the memory cost by truncating the expansion in (4) and storing only T most relevant historical observations. Once the buffer limit T is exceeded, the vector that is observed least recently is discarded to maintain a bound on the memory usage. Although SILK is used to approximate ILKM in [4], we prefer a small T in object tracking based on the observation that a relatively smaller T leads to more quick response to the changes while SILK with a larger T or the exact ILKM cause relatively longer delay.

III. Bayesian State Inference for Object Tracking

This letter utilizes the appearance model within the Bayesian inference framework to form an efficient tracking algorithm. Specifically, let $\mathbf{x}_t$ denote the state of a candidate at time t, which is characterized as a triple (x, y, scale) including the coordinates of the center and the ratio of the width of the bounding box of a candidate to that of the first frame. Given a set of observations $Z_t = \{\mathbf{z}_1, \ldots, \mathbf{z}_t\}$ up to time t ($\mathbf{z}_t$ is the concatenation vector of N patch descriptors), the posterior probability $p(\mathbf{x}_t|Z_t)$ is formulated by the Bayesian theorem as:

$$p(\mathbf{x}_t|Z_t) \propto p(Z_t|\mathbf{x}_t) \int p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}|Z_{t-1}) d\mathbf{x}_{t-1} \tag{9}$$

where $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ and $p(Z_t|\mathbf{x}_t)$ denote the dynamic model and the observation model respectively. In this letter, the dynamic model $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ is formulated by a Gaussian distribution with an assumption of *zero* acceleration:

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = N\left(\mathbf{x}_t|(\mathbf{x}_{t-1} + \nabla\mathbf{x}_{t-1}), \sum\right) \tag{10}$$

where $\nabla\mathbf{x}_{t-1}$ is the speed at time $t-1$ and Σ denotes a diagonal covariance matrix. A particle filter [9] is then used to approximate the distribution over the location of the object as well as the integral in (9). In addition, the observation model $p(Z_t|\mathbf{x}_t)$ is formulated as follows:

$$p(Z_t|\mathbf{x}_t) \propto F_t(\mathbf{z}_t) \tag{11}$$

where $F_t(\cdot)$ is the fused confidence defined in (6). Consequently maximum a posterior (MAP) estimation is carried out to find the best estimated candidate (i.e., $\hat{\mathbf{x}}_t$ and its corresponding $\hat{\mathbf{z}}_t$) of the *target*.

IV. Experiments

We evaluate the proposed method on four challenging sequences and compare it with two state-of-the-art algorithms, the covariance tracking (COVT) [16] and the multiple instance learning tracking (MILT) [2]. We implement COVT by ourselves and use the codes and settings of MILT from the website[2]. The sequences are collected from public dataset[3], internet and self-captured video, which are difficult due to occlusion, outliers, and appearance variation.

This letter utilizes the five-patch representation (i.e., $N = 5$) as illustrated in Fig. 1(a) and empirically sets the damping rate τ in (7), the update threshold tr, T for SILK, and the three standard deviations of the diagonal covariance matrix in (10) to 0.15, 3.78, 5 (consistent with the flexible setting of COVT in [16]), and (10, 10, 1.1) respectively. Since the states $\mathbf{x}$ is represented in a rectangle region, the calculation of Sigma set above can be accelerated through the integral image [18], which together with the Bayesian inference framework ensures a real time tracking system (about 10–15 frames per second on a desktop with Intel Core2 Duo CPU and 4 G memory). In addition, the linear combination coefficients $\alpha_t^{(n)}$ in (4) and the patch descriptors $\mathbf{v}^{(n)}$ are normalized respectively, to constrain the responses of ILKMs in (4) within a range of [0, 1].

Two measurements, Relative Position Errors (RPE) and the Percentage of Correctly tracked Frames (PCF) are computed to quantitatively evaluate tracking performance. Given the tracked region R_{tr} and the ground truth R_{gt}, RPE (as shown in Fig. 2) is defined as the ratio of the center distance between R_{tr} and R_{gt} to the size of R_{gt}. Moreover, PCF (as shown in Table I) measures the percentage of *correctly tracked* frames over all frames on the sequence. If and only if a ratio a_0 exceeds some value (45% here as a strict criterion), the corresponding frame is claimed *correctly tracked*. In particular, a_0 is defined as the size ratio

[2]http://vision.ucsd.edu/~bbabenko/project_miltrack.shtml.

[3]http://vision.stanford.edu/~birch/headtracker/seq/.

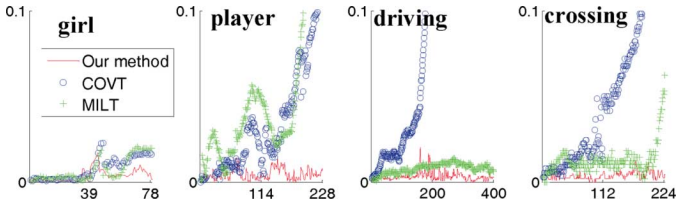

Fig. 2. Curves of relative position error (RPE) versus frame on four sequences. The red solid, blue circle, and green plus curves are for our method, COVT, and MILT respectively.

TABLE I
PERCENTAGE OF CORRECTLY TRACKED FRAMES (%) ON FOUR SEQUENCES

Seq.	Ours	MILT	COVT	Seq.	Ours	MILT	COVT
girl	*94*	72	76	*player*	*96*	09	69
driving	*98*	83	27	*crossing*	*100*	82	66

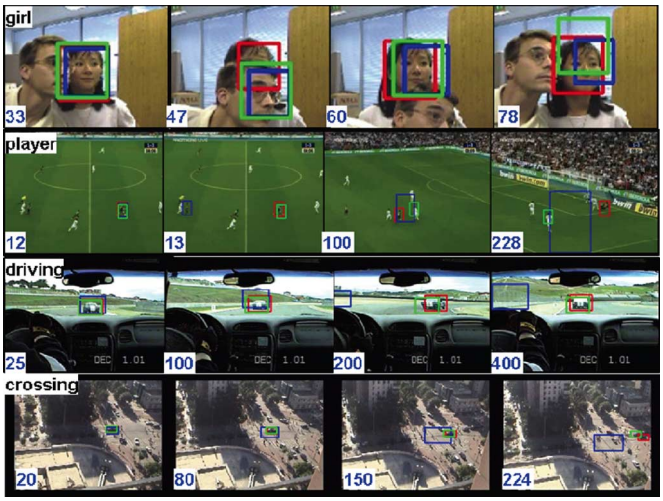

Fig. 3. Results on four sequences. Rows from top to bottom are for: *girl*, *player*, *driving*, and *crossing*. Red/blue/green rectangles are for our method, COVT, and MILT respectively. (Greyscale: Light/medium/dark gray rectangles are for MILT, our method, and COVT respectively.)

of the intersection to the union of R_{tr} and R_{gt}. We can easily observe that the proposed method outperforms both COVT [16] and MILT [2].

Some results are shown in Fig. 3. In the *girl* sequence, both COVT and MILT suffer from the occlusion caused by the man, while our method works well. In *player*, the player in black undergoes obvious appearance changes, which causes the failures of COVT and MILT in the 100th frame. COVT misses the target in the 13th frame as well. In *driving*, the vehicle changes in view and scale obviously. COVT begins to shift from the 100th frame. Although MILT can successfully track the car until the end of the sequence, its position errors are higher than ours. Similar cases happen in *crossing*, the illumination variation of the black car and outliers can be easily observed. COVT fails in the 150th frame whereas MILT is disturbed by another black car in the 224th frame. Nevertheless, the proposed method is capable of tracking the objects against all these difficulties.

V. CONCLUSION

This letter presents a Sigma set based implicit online learning method for visual tracking. We propose to utilize ILKM, which can effectively learn a robust predictor and promptly adapt to the appearance changes, to model each target patch. An arbitrator fusing the responses of patch ILKMs can handle partial occlusions and trigger the model update. Experimental results show that the proposed algorithm is more effective than both the covariance tracking and the multiple instance learning tracking algorithms on the challenging test sequences.

REFERENCES

[1] S. Avidan, "Ensemble tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 261–271, Feb. 2007.
[2] B. Babenko, M. Yang, and S. Belongie, "Visual tracking with online multiple instance learning," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2009.
[3] D. Brainard, "Color appearance and color difference specification," in *The Science of Color*, S. Shevell, Ed., 2nd ed. New York: Elsevier, 2003, pp. 202–206.
[4] L. Cheng, S. Vishwanathan, D. Schuurmans, S. Wang, and T. Caelli, "Implicit online learning with kernels," in *Proc. Advances in Neural Information Processing Systems*, 2007.
[5] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 564–577, May 2003.
[6] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," in *Proc. Eur. Conf. Computer Vision*, 2008.
[7] W. He, T. Yamashita, H. Lu, and S. Lao, "Surf tracking," in *Proc. IEEE Int. Conf. Computer Vision*, 2009.
[8] X. Hong, H. Chang, S. Shan, X. Chen, and W. Gao, "Sigma set: A small second order statistical region descriptor," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2009.
[9] M. Isard and A. Blake, "Condensation—conditional density propagation for visual tracking," *Int. J. Comput Vis.*, vol. 29, no. 1, pp. 5–28, 1998.
[10] X. Li, W. Hu, Z. Zhang, X. Zhang, M. Zhu, and J. Cheng, "Visual tracking via incremental log-Euclidean Riemannian subspace learning," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008.
[11] Y. Li, H. Ai, T. Yamashita, L. Lao, and M. Kawade, "Tracking in low frame rate video: A cascade particle filter with discriminative observers of different lifespans," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007.
[12] H. Lin and L. Li, "Support vector machinery for infinite ensemble learning," *J. Mach. Learn. Res.*, vol. 9, pp. 285–312, 2008.
[13] X. Liu and T. Yu, "Gradient feature selection for online boosting," in *Proc. IEEE Int. Conf. Computer Vision*, 2007.
[14] L. Lu and G. Hager, "A nonparametric treatment for location/segmentation based visual tracking," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007.
[15] I. Matthews, T. Ishikawa, and S. Baker, "The template update problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 810–815, Jun. 2004.
[16] F. Porikli, O. Tuzel, and P. Meer, "Covariance tracking using model update based on means on Riemannian manifolds," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2006.
[17] D. Ross, J. Lim, R. Lin, and M. Yang, "Incremental learning for robust visual tracking," *Int. J. Comput. Vis.*, vol. 77, no. 1–3, pp. 125–141, 2008.
[18] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," in *Proc. Eur. Conf. Computer Vision*, 2006.
[19] H. Wang, D. Suter, K. Schindler, and C. Shen, "Adaptive object tracking based on an effective appearance filter," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, pp. 1661–1667, Sep. 2007.
[20] Y. Wu, J. Cheng, J. Wang, and H. Lu, "Real-time visual tracking via incremental covariance tensor learning," in *Proc. IEEE Int. Conf. Computer Vision*, 2009.
[21] M. Yang, J. Yuan, and Y. Wu, "Spatial selection for attentional visual tracking," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007.
[22] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, 2006, Art. 13.