# Robust Visual Tracking with Double Bounding Box Model

Junseok Kwon[1], Junha Roh[2], Kyoung Mu Lee[3], and Luc Van Gool[1]

[1] Computer Vision Laboratory, ETH Zurich, Switzerland
[2] Imaging Media Research Center, KIST, Korea
[3] Department of ECE, ASRI, Seoul National University, Korea
{kwonj,vangool}@vision.ee.ethz.ch, junha.roh@imrc.kist.re.kr, kyoungmu@snu.ac.kr

**Abstract.** A novel tracking algorithm that can track a highly non-rigid target robustly is proposed using a new bounding box representation called the Double Bounding Box (DBB). In the DBB, a target is described by the combination of the Inner Bounding Box (IBB) and the Outer Bounding Box (OBB). Then our objective of visual tracking is changed to find the IBB and OBB instead of a single bounding box, where the IBB and OBB can be easily obtained by the Dempster-Shafer (DS) theory. If the target is highly non-rigid, any single bounding box cannot include all foreground regions while excluding all background regions. Using the DBB, our method does not directly handle the ambiguous regions, which include both the foreground and background regions. Hence, it can solve the inherent ambiguity of the single bounding box representation and thus can track highly non-rigid targets robustly. Our method finally finds the best state of the target using a new Constrained Markov Chain Monte Carlo (CMCMC)-based sampling method with the constraint that the OBB should include the IBB. Experimental results show that our method tracks non-rigid targets accurately and robustly, and outperforms state-of-the-art methods.

## 1 Introduction

Visual tracking has been used in many artificial intelligence applications, including surveillance, augmented reality, medical imaging, and other intelligent vision systems [5,8,16,32,34]. In practical applications, the purpose of visual tracking is to find the best configuration of a target with a given observation [38]. As many conventional tracking methods describe the target with a single bounding box, the configuration at time $t$ is typically represented by a three-dimensional vector $\mathbf{X}_t = (X_t^x, X_t^y, X_t^s)$, where $(X_t^x, X_t^y)$, and $X_t^s$ are the center coordinate and the scale of the target, respectively. This single bounding box representation is widely used because it allows the easy inference of the best configuration using a low-dimensional vector [1,4,13,17,20,24]. In addition, with the representation, tracking methods that use the tracking by detection approach [3,7,14,18,31,35] easily train a classifier using the rectangular patches described by the bounding box.

D. Fleet et al. (Eds.): ECCV 2014, Part I, LNCS 8689, pp. 377–392, 2014.

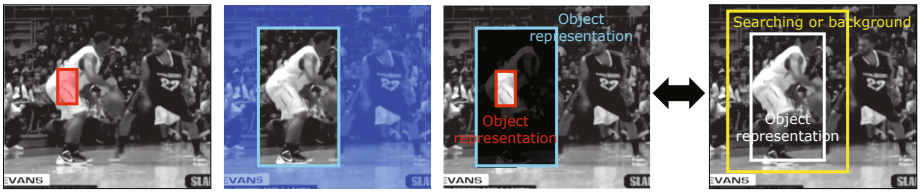



**Fig. 1. Example of the different bounding box representations**. (a) The IBB is the bounding box that only includes the target region, but excludes some parts of it. The OBB is the bounding box that includes whole target region but also includes some background regions. In our DBB, the bounding box of the target is represented by the combination of the IBB (red) and OBB (blue). (b) The discriminative tracking approaches in [2][36] also use two BBs for visual tracking. However only one BB is used for representing the target configuration. The other BB is required only to get the background information [2] or to make a search range [36], which is quite different from our OBB. All of two BBs in our method are designed to represent the non-rigid target configuration.

However, the single bounding box representation has inherent ambiguity if a target is highly non-rigid. No single bounding box can cover the whole region of the target while excluding all the background regions, as shown in Fig.1(a). In this case, the tracking methods may choose the Inner Bounding Box (IBB), the Outer Bounding Box (OBB), or between them. The IBB is the largest bounding box that contains pure object region only. The OBB is the smallest bounding box that contains the whole object region, where the outside of the OBB is pure background. However, although the IBB in Fig.1(a) can deliver the pure target region, the IBB and bounding boxes that are smaller than the IBB lose lots of useful information about the target by excluding large parts of it. While the OBB in Fig.1(a) includes the whole target region, the OBB and bounding boxes that are larger than the OBB inevitably contain unwanted background regions. The bounding boxes between the IBB and OBB include both the target region and the background region. Now, the natural question is *which box representation can best describe a general and non-rigid target*. The present paper aims to resolve this ambiguity in the bounding box representation, and to track a non-rigid target robustly using a new bounding box representation.

In this paper, the bounding box of a target is represented by the combination of the IBB and OBB. We call this representation the DBB. By describing the bounding box with this combination, our method solves the inherent ambiguity in the single bounding box representation. In addition, this representation improves tracking performance because it does not consider the ambiguous regions (the black region between the IBB and OBB in Fig.1(a)) in determining the probable configuration of the target, which contains mixed target and background regions. Instead, our method only considers the *maximal pure foreground*

*region*, inside of the IBB (the red region in Fig.1(a)) and the *pure background region*, outside of the OBB (the blue region in Fig.1(a)) in determining the probable configuration of the target. Our method finds the IBB and OBB using the DS theory [9,11] in Section 3.2. Following the DS theory, the IBB is obtained by maximizing the similarity between the red region and the target model, and the OBB is obtained by maximizing the dissimilarity between the green region and the target model. With the IBB and OBB, the method searches the best state of the target using the proposed Constrained Maximum a Posteriori (CMAP) estimate in Section 4.1. The best state maximizes the posterior probability while it satisfies the constraint that the OBB has to include the IBB. The best state can be achieved by the proposed CMCMC-based sampling method in Section 4.2.

The first contribution of our work is to propose a new bounding box representation. A highly non-rigid object cannot be adequately described by any single bounding box. To solve the ambiguity in the conventional bounding box representation, the target is represented by the combination of two bounding boxes, which is called the DBB. The second contribution is to apply the the DS theory to the bounding box representation problem and to provide a theoretical basis for determining the IBB and OBB. The last contribution is to present an efficient tracking system using the DBB and a new CMCMC-based sampling method. The DBB explores the complementary connection between the IBB and OBB to track highly non-rigid targets accurately. In practice, the IBB is robust to the deformation of the target but sensitive to noise because of its small size. The OBB is resistant to noise but imprecise on the deformation of the target because of its large size. Hence, these two representations complement each other, resulting in a representation that is insensitive to both deformation and noise. The CMCMC-based sampling method efficiently determines the best states of the target while maintaining the constraint in which the OBB must include the IBB.

## 2 Related Work

In tracking methods for non-rigid targets, the BHT tracker [26] described the target using multiple rectangular blocks, whose positions within the tracking window are adaptively determined. The BHMC tracker [22] represented the target using multiple local patches, in which the topology among local patches continuously evolves. Using multiple rectangular blocks and multiple local patches, these methods efficiently track non-rigid objects undergoing large variations in appearance and shape. Our method and aforementioned part models are similar in the sense of using multiple bounding boxes. However, the main advantages of our method over these methods are in how it determines the sizes of the IBB and OBB and how to calculate their likelihoods efficiently using the DS theory [9,11]. Therefore, our method can improve even the conventional part models by applying the aforementioned two contributions to them. In addition, compared with the part models [29], our method describes the target as a low-dimensional vector using only the IBB and OBB. Hence, the best state of the target is easily determined with a much smaller computational cost.

In sampling-based tracking methods, the particle filter [15] determines the best state by considering the non-Gaussian and the multi-modality of the target distribution in the tracking problem. Markov Chain Monte Carlo (MCMC) methods [19,33] efficiently determines the best state in high-dimensional state spaces. Compared with these methods, the CMCMC-based sampling method finds the best state while satisfying some constraints. It is a more difficult problem than the conventional sampling problems.

In tracking methods using the the DS theory [9,11], the method in [10] presented a face tracking system using a pixel fusion process from three color sources within the framework of the DS theory. The methods in [23,25] resolved the visual tracking problem by combining evidences from multiple sources using the DS theory. Unlike these methods that used the DS theory only for the observation fusion, our method employs the theory for the bounding box representation.

## 3 Design of the Double Bounding Box

### 3.1 Bounding Box Representation Using DS Theory

In the tracking method, a state $\mathbf{X}_t$ represents a bounding box of a rectangular region at $t$-th frame. Let us denote $\mathbb{R}(\mathbf{X}_t)$ as the region enclosed by $\mathbf{X}_t$. According to the Shafer's framework [9], we can define a mass function $m(\mathbb{R}(\mathbf{X}_t))$ for the region of a bounding box $\mathbb{R}(\mathbf{X}_t)$, which is bounded by two values, i.e., belief and plausibility:

$$bel(\mathbb{R}(\mathbf{X}_t)) \leq m(\mathbb{R}(\mathbf{X}_t)) \leq pl(\mathbb{R}(\mathbf{X}_t)), \tag{1}$$

where $bel(\mathbb{R}(\mathbf{X}_t))$ and $pl(\mathbb{R}(\mathbf{X}_t))$ denote the belief and the plausibility of $\mathbb{R}(\mathbf{X}_t)$, respectively. In this work, the mass function $m(\mathbb{R}(\mathbf{X}_t))$ is designed by the likelihood of the region of the bounding box:

$$m(\mathbb{R}(\mathbf{X}_t)) \equiv p(\mathbf{Y}_t | R = \mathbb{R}(\mathbf{X}_t)) = \frac{1}{c} e^{-dist\left(M_t, \mathbf{Y}_t(\mathbb{R}(\mathbf{X}_t))\right)}, \tag{2}$$

where $\mathbf{Y}_t(\mathbb{R}(\mathbf{X}_t))$ denotes the observation inside of the region $\mathbb{R}(\mathbf{X}_t)$ and $dist(\cdot)$ returns the distance between the observation $\mathbf{Y}_t(\mathbb{R}(\mathbf{X}_t))$ and the target model $M_t$. For the observation and the distance measure, we utilize the HSV color histogram and the Bhattacharyya similarity coefficient in [28]. $c$ in (2) is a normalization constant.

Now, given an IBB $\mathbf{X}_t^i$ and an OBB $\mathbf{X}_t^o$, as shown in Fig.2, the whole region of interest can be divided by three regions $\mathbf{R}_t^1$, $\mathbf{R}_t^2$, and $\mathbf{R}_t^3$, making the frame of discernment to be $\mathbf{U}_t = \{\mathbf{R}_t^1, \mathbf{R}_t^2, \mathbf{R}_t^3\}$, where the regions are mutually exclusive $\bigcap_{i=1}^3 \mathbf{R}_t^i = \phi$ and the union of the regions compose a whole region $\bigcup_{i=1}^3 \mathbf{R}_t^i = \mathbf{U}_t$. Note that the IBB $\mathbf{X}_t^i$ represents the region $\mathbf{R}_t^1$, and the OBB $\mathbf{X}_t^o$ covers the region $\{\mathbf{R}_t^1, \mathbf{R}_t^2\}$, where $\{\mathbf{R}_t^1, \mathbf{R}_t^2\}$ denotes a union of regions $\mathbf{R}_t^1$ and $\mathbf{R}_t^2$. The power set of the universal set, $2^{\mathbf{U}_t}$, is $\{\phi,\ \mathbf{R}_t^1,\ \mathbf{R}_t^2,\ \mathbf{R}_t^3,\ \{\mathbf{R}_t^1, \mathbf{R}_t^2\},\ \{\mathbf{R}_t^1, \mathbf{R}_t^3\},\ \{\mathbf{R}_t^2, \mathbf{R}_t^3\},\ \mathbf{U}_t\}$. According to the DS theory [9,11], the mass function in (2) is then normalized, such that the masses of the elements of the power set $2^{\mathbf{U}_t}$ add up to a total of 1. In this paper, the mass corresponds to the likelihood score

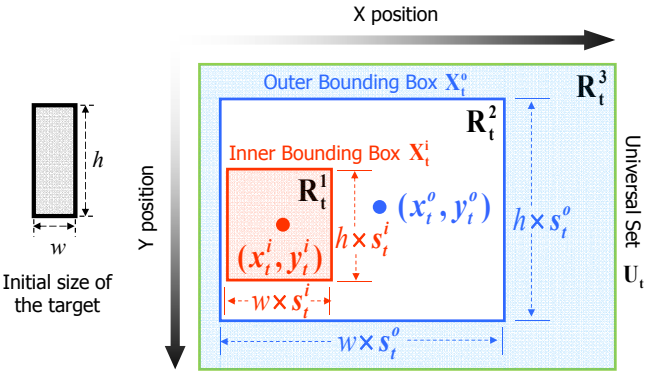



**Fig. 2. Notation of the DBB**. $\mathbf{R}_t^1$ indicates the region inside of the IBB $\mathbf{X}_t^i$. $\mathbf{R}_t^3$ indicates the region outside of the OBB $\mathbf{X}_t^o$. $\mathbf{R}_t^2$ indicates the region between the IBB and OBB.

$p(\mathbf{Y}_t|R)$ and, thus, we make the sum of the likelihood scores of all elements of $2^{\mathbf{U}_t}$ to be 1:

$$\sum_{r|r\in 2^{\mathbf{U}_t}} p(\mathbf{Y}_t|R=r) = p(\mathbf{Y}_t|R=\mathbf{R}_t^1) + p(\mathbf{Y}_t|R=\mathbf{R}_t^2) + p(\mathbf{Y}_t|R=\mathbf{R}_t^3) + p(\mathbf{Y}_t|R=\{\mathbf{R}_t^1,\mathbf{R}_t^2\}) + p(\mathbf{Y}_t|R=\{\mathbf{R}_t^2,\mathbf{R}_t^3\}) = 1. \tag{3}$$

In (3), $p(\mathbf{Y}_t|R=\phi)$, $p(\mathbf{Y}_t|R=\mathbf{U}_t)$, and $p(\mathbf{Y}_t|R=\{\mathbf{R}_t^1,\mathbf{R}_t^3\})$ are designed as zero because $\phi$, $\mathbf{U}_t$, and $\{\mathbf{R}_t^1,\mathbf{R}_t^3\}$ make meaningless regions, which are empty region, entire region, and two separate regions, respectively.

### 3.2 Inner and Outer Bounding Boxes

According to the DS theory [9,11], the belief in (1) is defined as the sum of all the masses of subsets of the set of interest. In our problem, the mass corresponds to the likelihood score $p(\mathbf{Y}_t|R)$, whereas the set of interest is the IBB, $\mathbb{R}(\mathbf{X}_t^i) = \{\mathbf{R}_t^1\}$, or OBB, $\mathbb{R}(\mathbf{X}_t^o) = \{\mathbf{R}_t^1,\mathbf{R}_t^2\}$. The belief of the IBB is then modeled by the sum of the likelihood scores of all subsets of $\{\mathbf{R}_t^1\}$, as follows:

$$bel(\mathbb{R}(\mathbf{X}_t^i)) = p(\mathbf{Y}_t|R=\mathbf{R}_t^1). \tag{4}$$

The belief of the OBB is modeled by the sum of the likelihood scores of all subsets of $\{\mathbf{R}_t^1,\mathbf{R}_t^2\}$, as follows:

$$bel(\mathbb{R}(\mathbf{X}_t^o)) = p(\mathbf{Y}_t|R=\mathbf{R}_t^1) + p(\mathbf{Y}_t|R=\mathbf{R}_t^2) + p(\mathbf{Y}_t|R=\{\mathbf{R}_t^1,\mathbf{R}_t^2\}). \tag{5}$$

Notably, the belief is the amount of belief that directly supports the bounding box at least in part, forming a lower bound [9].

The plausibility in (1) is sum of all the masses of subsets that intersect the set of interest. Thus the plausibility of the IBB is designed as the sum of the likelihood scores of all subsets of $\mathbf{U}_t$, which intersect $\mathbb{R}(\mathbf{X}_t^i) = \{\mathbf{R}_t^1\}$.

$$pl(\mathbb{R}(\mathbf{X}_t^i)) = p(\mathbf{Y}_t|R = \mathbf{R}_t^1) + p(\mathbf{Y}_t|R = \{\mathbf{R}_t^1, \mathbf{R}_t^2\}), \tag{6}$$

where $p(\mathbf{Y}_t|R = \mathbf{U}_t) = p(\mathbf{Y}_t|R = \{\mathbf{R}_t^1, \mathbf{R}_t^3\}) = 0$. Because $p(\mathbf{Y}_t|R = \{\mathbf{R}_t^1, \mathbf{R}_t^2\})$ is positive and, thus, $pl(\mathbb{R}(\mathbf{X}_t^i))$ in (6) is larger than $bel(\mathbb{R}(\mathbf{X}_t^i))$ in (4), the IBB satisfies (1): $bel(\mathbb{R}(\mathbf{X}_t^i)) \leq m(\mathbb{R}(\mathbf{X}_t^i)) \leq pl(\mathbb{R}(\mathbf{X}_t^i))$. The plausibility of the OBB is designed as the sum of the likelihood scores of all subsets of $\mathbf{U}_t$, which intersect $\mathbb{R}(\mathbf{X}_t^o) = \{\mathbf{R}_t^1, \mathbf{R}_t^2\}$.

$$\begin{aligned} pl(\mathbb{R}(\mathbf{X}_t^o)) = p(\mathbf{Y}_t|R = \mathbf{R}_t^1) + p(\mathbf{Y}_t|R = \mathbf{R}_t^2) + p(\mathbf{Y}_t|R = \{\mathbf{R}_t^1, \mathbf{R}_t^2\}) \\ + p(\mathbf{Y}_t|R = \{\mathbf{R}_t^2, \mathbf{R}_t^3\}) = 1 - p(\mathbf{Y}_t|R = \mathbf{R}_t^3), \end{aligned} \tag{7}$$

where the last equality holds because of (3). Because $p(\mathbf{Y}_t|R = \{\mathbf{R}_t^2, \mathbf{R}_t^3\})$ is positive and, thus, $pl(\mathbb{R}(\mathbf{X}_t^o))$ in (7) is larger than $bel(\mathbb{R}(\mathbf{X}_t^o))$ in (5), the OBB satisfies (1): $bel(\mathbb{R}(\mathbf{X}_t^o)) \leq m(\mathbb{R}(\mathbf{X}_t^o)) \leq pl(\mathbb{R}(\mathbf{X}_t^o))$. Note that the plausibility forms an upper bound because there is only so much evidence which contradicts that bounding box [9].

Then, with (4),(6),(5), and (7), we obtain

$$\begin{aligned} bel(\mathbb{R}(\mathbf{X}_t^i)) \leq pl(\mathbb{R}(\mathbf{X}_t^i)) \leq bel(\mathbb{R}(\mathbf{X}_t^o)) \leq pl(\mathbb{R}(\mathbf{X}_t^o)), \\ bel(\mathbb{R}(\mathbf{X}_t^i)) \leq m(\mathbb{R}(\mathbf{X}_t^i)), m(\mathbb{R}(\mathbf{X}_t^o)) \leq pl(\mathbb{R}(\mathbf{X}_t^o)). \end{aligned} \tag{8}$$

In terms of the DS theory, the best bounding box has the largest belief and plausibility values. Thus, (8) is maximized to obtain the best IBB and OBB: $\max bel(\mathbb{R}(\mathbf{X}_t^i)) \leq \max m(\mathbb{R}(\mathbf{X}_t^i)), \max m(\mathbb{R}(\mathbf{X}_t^o)) \leq \max pl(\mathbb{R}(\mathbf{X}_t^o))$. Thereafter, the best IBB $\hat{\mathbf{X}}_t^i$ is obtained using (4):

$$\hat{\mathbf{X}}_t^i = \underset{\mathbf{X}_t^i}{\arg\max}\, bel(\mathbb{R}(\mathbf{X}_t^i)) = \underset{\mathbf{X}_t^i}{\arg\max}\, p(\mathbf{Y}_t|R = \mathbf{R}_t^1), \tag{9}$$

where the best IBB $\hat{\mathbf{X}}_t^i$ is determined by maximizing the similarity between the whole region inside the IBB, $\mathbf{R}_t^1$, and the target model $M_t$. Similarly, the best OBB $\hat{\mathbf{X}}_t^o$ is obtained using (7):

$$\hat{\mathbf{X}}_t^o = \underset{\mathbf{X}_t^o}{\arg\max}\, pl(\mathbb{R}(\mathbf{X}_t^o)) = \underset{\mathbf{X}_t^o}{\arg\max} \left[1 - p(\mathbf{Y}_t|R = \mathbf{R}_t^3)\right], \tag{10}$$

where the OBB $\hat{\mathbf{X}}_t^o$ is determined by maximizing the dissimilarity between the region outside of the OBB, $\mathbf{R}_t^3$, and the target model $M_t$.

## 4 Visual Tracker Using the Double Bounding Box

### 4.1 Constrained Maximum a Posteriori

As illustrated in Fig.2, our state $\mathbf{X}_t^{DBB} = (\mathbf{X}_t^i, \mathbf{X}_t^o)$ is represented as the combination of the sub-states of the IBB and OBB. $\mathbf{X}_t^i$ and $\mathbf{X}_t^o$ consist of a three-dimensional vector including the center coordinate and the scale of the IBB and

OBB, respectively. Thus, the cardinality of the final state $\mathbf{X}_t^{DBB}$ is 6. Then, the objective of our tracking problem is to find the best state $\hat{\mathbf{X}}_t^{DBB} = (\hat{\mathbf{X}}_t^i, \hat{\mathbf{X}}_t^o)$ that maximizes the posterior $p(\mathbf{X}_t^{DBB}|\mathbf{Y}_{1:t})$:

$$\hat{\mathbf{X}}_t^{DBB} = \underset{\mathbf{X}_t^i, \mathbf{X}_t^o}{\arg\max}\, p(\mathbf{X}_t^i, \mathbf{X}_t^o|\mathbf{Y}_{1:t}) \; subject\ to \;\; \mathbb{R}(\mathbf{X}_t^i) \subset \alpha\mathbb{R}(\mathbf{X}_t^o) \subset \beta\mathbb{R}(\mathbf{X}_t^i), \tag{11}$$

where $\alpha\mathbb{R}(\mathbf{X}_t^o)$ means that the width and height of the region $\mathbb{R}(\mathbf{X}_t^o)$ are multiplied by $\alpha$. $\alpha > 1$ practically makes (11) to be easily solved by relaxing the strong constraint, $\alpha = 1$, although we get the approximated solution. Because $\alpha > 1$, a small part of the IBB can be located outside of the OBB in the experimental results. In (11), $p(\mathbf{X}_t^i, \mathbf{X}_t^o|\mathbf{Y}_{1:t})$ is reformulated by

$$\begin{aligned} p(\mathbf{X}_t^i, \mathbf{X}_t^o|\mathbf{Y}_{1:t}) \propto\, & p(\mathbf{Y}_t|\mathbf{X}_t^i)p(\mathbf{Y}_t|\mathbf{X}_t^o)\times \\ & \int p(\mathbf{X}_t^i, \mathbf{X}_t^o|\mathbf{X}_{t-1}^i, \mathbf{X}_{t-1}^o)p(\mathbf{X}_{t-1}^i, \mathbf{X}_{t-1}^o|\mathbf{Y}_{1:t-1})d\mathbf{X}_{t-1}, \end{aligned} \tag{12}$$

where $p(\mathbf{Y}_t|\mathbf{X}_t^i)$ is the likelihood of the IBB and $p(\mathbf{Y}_t|\mathbf{X}_t^o)$ is the likelihood of the OBB.

In (12), we design $p(\mathbf{Y}_t|\mathbf{X}_t^i)$ to measure the similarity between the region inside of the IBB, $\mathbf{R}_t^1$, and the target model, $M_t$, based on (2) and (9) derived from the DS theory:

$$p(\mathbf{Y}_t|\mathbf{X}_t^i) \equiv p(\mathbf{Y}_t|R = \mathbf{R}_t^1) = \frac{1}{c} e^{-dist\left(M_t, \mathbf{Y}_t(\mathbf{R}_t^1)\right)}, \tag{13}$$

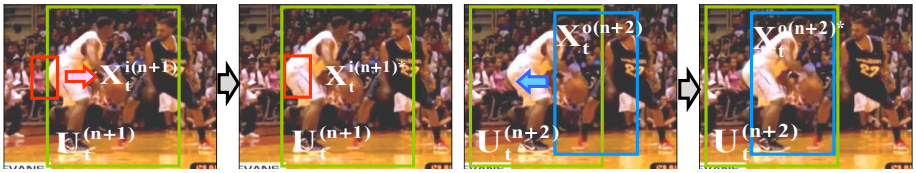
where $c$ is a normalization constant. We design $p(\mathbf{Y}_t|\mathbf{X}_t^o)$ to measure the dissimilarity between the region outside of the OBB, $\mathbf{R}_t^3$, and the target model, $M_t$, based on (2) and (10) derived from the DS theory:

$$p(\mathbf{Y}_t|\mathbf{X}_t^o) \equiv 1 - p(\mathbf{Y}_t|R = \mathbf{R}_t^3) = 1 - \frac{1}{c} e^{-dist\left(M_t, \mathbf{Y}_t(\mathbf{R}_t^3)\right)}. \tag{14}$$

Note that the target model, $M_t$, is updated over time by averaging the initial model with the most recent model.

In (12), the dynamical part, $p(\mathbf{X}_t^i, \mathbf{X}_t^o|\mathbf{X}_{t-1}^i, \mathbf{X}_{t-1}^o)$, is realized by proposal and constraint steps of the CMCMC, explained in the next section, where a new IBB(OBB) is proposed based on the previous OBB(IBB) and a new IBB is included by a new OBB.

The first constraint in (11) makes the region described by the OBB, $\mathbb{R}(\mathbf{X}_t^o)$, to include the region described by the IBB, $\mathbb{R}(\mathbf{X}_t^i)$, while the second constraint prevents the OBB from becoming infinitely large. Compared with MAP, our CMAP in (11) is more difficult because it should satisfy the aforementioned constraints. To obtain the best state, searching all states within the state space is impractical. Thus, our method adopts the MCMC sampling method [19], which produces $N$ number of sampled states. Among the sampled states, the sampling method easily chooses the best one that maximizes the posterior probability in (11). We modify the MCMC sampling method to satisfy the aforementioned constraint and present a new CMCMC-based sampling method, which will be explained in the next section.




(a) Modification of the IBB (b) Modification of the OBB

**Fig. 3.** Example of the DBB constraint

### 4.2 Constrained Markov Chain Monte Carlo

The CMCMC-based sampling method defines a single Markov Chain and obtains samples over the chain. As we define two sub-states, $\mathbf{X}_t^i$ and $\mathbf{X}_t^o$, we get samples of the sub-states, alternately. For example, we get samples like $\mathbf{X}_t^{i(n-1)}$, $\mathbf{X}_t^{o(n)}$, $\mathbf{X}_t^{i(n+1)}$,..., where $\mathbf{X}_t^{o(n)}$ and $\mathbf{X}_t^{i(n+1)}$ are the $n$-th sample for the OBB and the *(n+1)*-th sample for the IBB, respectively.

• **Proposal Step of the IBB** First, the method obtains a sample of the IBB $\mathbf{X}_t^i$ by three main steps: the proposal, constraint, and acceptance steps. The proposal step suggests a new sample of the IBB as follows: a new center of the IBB, $c(\mathbf{X}_t^{i(n+1)})$, is proposed through the Gaussian function $G$ (mean: current center of the OBB, $c(\mathbf{X}_t^{o(n)})$, variance: 0.5). A new scale of the IBB, $s(\mathbf{X}_t^{i(n+1)})$, is proposed through $G$ (mean: current scale of the IBB, $s(\mathbf{X}_t^{i(n-1)})$, variance: 0.01).

• **Constraint Step of the IBB** To satisfy the constraint in (11), the bounding box of the proposed sample $\mathbf{X}_t^{i(n+1)}$ shifts to the bounding box of the sample $\mathbf{X}_t^{i(n+1)^*}$, where the shifted bounding box is included by the region $\alpha\mathbb{R}(\mathbf{X}_t^{o(n)})$, as shown in Fig.3(a): $\mathbb{R}(\mathbf{X}_t^{i(n+1)^*}) \subset \alpha\mathbb{R}(\mathbf{X}_t^{o(n)}) = \mathbb{R}(\mathbf{U}_t^{(n+1)})$, where $\mathbf{U}_t^{(n+1)}$ is the *(n+1)*-th universal set. The region $\alpha\mathbb{R}(\mathbf{X}_t^{o(n)})$ has the same center as $\mathbf{X}_t^{o(n)}$ and is $\alpha$ times the scale of $\mathbf{X}_t^{o(n)}$, where $\alpha$ is empirically set to 1.2.

• **Acceptance Step of the IBB** After the proposed step, the acceptance step determines whether the proposed sample $\mathbf{X}_t^{i(n+1)^*}$ is accepted or not with the probability, $\min\left[1, \frac{p(\mathbf{Y}_t|\mathbf{X}_t^{i(n+1)^*})}{p(\mathbf{Y}_t|\mathbf{X}_t^{i(n-1)})}\right]$, where $p(\mathbf{Y}_t|\mathbf{X}_t^{i(n+1)^*})$ and $p(\mathbf{Y}_t|\mathbf{X}_t^{i(n-1)})$ are the likelihoods of the proposed and current IBB, respectively. The likelihood of the IBB is defined in (13).

• **Proposal Step of the OBB** Our method then proposes a sample of the OBB $\mathbf{X}_t^o$ as follows: a new center of the OBB, $c(\mathbf{X}_t^{o(n+2)})$, is proposed through $G$ (mean: current center of the IBB, $c(\mathbf{X}_t^{i(n+1)})$, variance: 0.5). A new scale of the OBB, $s(\mathbf{X}_t^{o(n+2)})$, is proposed through $G$ (mean: current scale of the OBB, $s(\mathbf{X}_t^{o(n)})$, variance: 0.01).

• **Constraint Step of the OBB** To satisfy the constraint in (11), the bounding box of the proposed sample $\mathbf{X}_t^{o(n+2)}$ shifts to the bounding box of the sample $\mathbf{X}_t^{o(n+2)^*}$, where the shifted bounding box is included by the region $\beta\mathbb{R}(\mathbf{X}_t^{i(n+1)})$,


(a) IBB only (b) DBB (c) OBB only (d) DBB

**Fig. 4. Performance of the DBB** in *basketball* seq. which has abrupt motions and pose variations. The red and blue rectangles are the IBB and OBB, respectively.

as shown in Fig.3(b): $\mathbb{R}(\mathbf{X}_t^{o(n+2)^*}) \subset \beta\mathbb{R}(\mathbf{X}_t^{i(n+1)}) = \mathbb{R}(\mathbf{U}_t^{(n+2)})$, where $\mathbf{U}_t^{(n+2)}$ is the *(n+2)*-th universal set. The region $\beta\mathbb{R}(\mathbf{X}_t^{i(n+1)})$ has the same center as $\mathbf{X}_t^{i(n+1)}$ and is $\beta$ times the scale of $\mathbf{X}_t^{i(n+1)}$, where $\beta$ is empirically set to 5.0.

• **Acceptance Step of the OBB** After the proposal step, the method accepts the proposed sample $\mathbf{X}_t^{o(n+2)^*}$ with the probability, $\min\left[1, \frac{p(\mathbf{Y}_t|\mathbf{X}_t^{o(n+2)^*})}{p(\mathbf{Y}_t|\mathbf{X}_t^{o(n)})}\right]$, where $p(\mathbf{Y}_t|\mathbf{X}_t^{o(n+2)^*})$ and $p(\mathbf{Y}_t|\mathbf{X}_t^{o(n)})$ are the likelihoods of the proposed and current OBB, respectively. The likelihood of the OBB is defined in (14). These steps iteratively continue until the number of iterations reaches the predefined value.

## 5 Experimental Results

To initialize the proposed method (DBB), we manually draw the IBB and OBB at the first frame. The IBB and OBB could have different width/height proportion initially. However, the proportion is fixed for each bounding box during the tracking process. The number of samples was fixed to 1000 for all sampling-based methods, including our method. For all experiments, we used the fixed parameters. Our method approximately takes 0.1 sec per frame.

### 5.1 Analysis of the Proposed Method

**Analysis of the DBB:** The performance difference between the single bounding box representation and the DBB representation were examined. The experiments were performed under the same conditions, differing only in the types of bounding box representation. As shown in Fig.4, either the IBB alone or the OBB alone is prone to drift away from the target. Fig.4(a) shows that the IBB began to drift and to track the background region around the target, as the appearance of the target became severely deformed. Our method kept tracking the target because of the constraint provided by the OBB. Thus, the OBB serves as a weak constraint that gives an estimate of the position of the target to the IBB, as it began sampling from the position of the IBB of the previous frame. If the IBB includes some parts of the background, then it will have a tendency

**Table 1. Comparison of tracking results** using IMCMC and CMCMC. The numbers indicate the average center location errors in pixels. These numbers were obtained by running each algorithm five times and averaging the results.

| | *basketball* | *lazysong* | *fx* | *diving* | *gymnastics* | *faceocc* | *twinings* | *singer1* | *skating2* | **Average** |
|---|---|---|---|---|---|---|---|---|---|---|
| IMCMC | 209 | 42 | 44 | 56 | 109 | 21 | 18 | 84 | 56 | **71.0** |
| CMCMC | 36 | 17 | 25 | 16 | 16 | 6 | 6 | 12 | 28 | **18.0** |


(a) Sensitive to initialization (b) Sensitive to parameters

**Fig. 5.** Stability of our method

to drift because it recognizes the background part as the target. However, in the DBB, the IBB is sampled at the estimated position from the OBB, pulling the IBB to the center of the target and reducing the possibility of drifting. The OBB can estimate the target position better than the IBB, owing to its large region, which makes the OBB robust to noise. Similarly, the IBB also helps the OBB. In Fig.4(c), the OBB drifted, despite being insensitive to noise. As in the previous case, the IBB complemented the OBB, as shown in Fig.4(d). The IBB usually has a higher probability of only including the target than the OBB. Thus, it also serves as a weak constraint, pulling the OBB to the target region.

**Analysis of the CMCMC:** The performances of the Interacting Markov Chain Monte Carlo (IMCMC) in [20] and CMCMC were also compared to demonstrate the superiority of the proposed CMCMC. The fundamental problem of IMCMC is that IMCMC has no mechanism to satisfy the constraint. Hence, we can't use IMCMC for our problem. Although we forcibly make IMCMC to satisfy the constraint, IMCMC is experimentally worse than CMCMC. For this experiment, IMCMC was applied to incorporate purposely the same capability of pulling one box to another, similar to CMCMC. IMCMC initially verifies the constraint in which the IBB must be inside the OBB. If this is verified, it separately samples each Markov Chain for each bounding box; otherwise, it provides an offset to one of the bounding boxes and probabilistically determines which one has to be adjusted. However, it cannot prevent itself from drifting away even if the two bounding boxes do satisfy the constraint. When the IBB begins to drift, it forces the OBB to drift as well because of the constraint, rendering the IBB even worse. As shown in Table 1, IMCMC cannot outperform CMCMC, because the probability that the IBB will pull the OBB toward it and the vulnerability of the IBB to noise are high.

**Table 2. Quantitative comparison of tracking results** with other methods. In this experiment, other tracking methods utilize the **IBB representation**. The numbers indicate the average center location errors in pixels and the amount of successfully tracked frames (score $> 0.5$), where the score is defined by the overlap ratio between the predicted bounding box $B_p$ and the ground truth bounding box $B_{gt}$: $\frac{area(B_p \bigcap B_{gt})}{area(B_p \bigcup B_{gt})}$. The best result is shown in red and the second-best in blue. N/W means that a method does not work at the corresponding dataset. For our method, the mean of center positions and bounding boxes of the IBB and OBB are reported as the final tracking result.

| | ABCShift | BHMC | BHT | MIL | IVT | VTD | VTS | MC | HT | LGT | TLD | DBB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *basketball* | 80/33 | 80/34 | 63/40 | 133/22 | 50/49 | 58/42 | **38**/**65** | 110/26 | 197/18 | 160/20 | 178/18 | **36**/**65** |
| *lazysong* | 71/40 | 55/50 | 142/39 | 38/55 | 38/55 | **17**/**70** | 26/67 | 30/62 | 56/50 | 42/55 | 20/68 | **17**/**69** |
| *fx* | 59/49 | 73/39 | 69/41 | 56/49 | 46/54 | 33/65 | **29**/**69** | 144/19 | 70/41 | 126/21 | 30/65 | **25**/**70** |
| *diving* | 23/35 | 41/39 | N/W | 76/26 | 68/31 | 23/34 | **16**/**46** | 20/34 | 76/26 | **15**/**51** | N/W | **16**/**46** |
| *gymnastics* | 45/45 | 29/59 | N/W | 42/47 | 62/41 | 22/62 | 18/66 | 17/66 | 108/23 | 99/31 | **13**/**72** | **16**/**68** |
| *faceocc* | 29/70 | 50/50 | N/W | 36/69 | 61/44 | 21/75 | 20/**79** | 20/78 | 34/70 | **19**/78 | 25/72 | **6**/**88** |
| *twinings* | 18/76 | **5**/**91** | 34/67 | 15/80 | 17/77 | 9/87 | 8/87 | 14/81 | 31/68 | 22/75 | 15/80 | **6**/**89** |
| *tiger1* | 85/30 | 33/30 | 30/45 | 17/59 | 80/30 | 16/61 | 15/62 | 30/42 | 30/42 | 15/62 | **9**/**70** | **13**/**65** |
| *david* | 50/21 | 40/31 | 17/50 | 29/47 | 13/58 | 13/58 | 10/60 | 42/30 | 10/60 | **6**/**71** | 8/62 | **6**/64 |
| *shaking* | 50/55 | 47/60 | 22/72 | 27/70 | 107/22 | 8/88 | 7/89 | 99/21 | 15/76 | 15/76 | **5**/**90** | **6**/**90** |
| *soccer* | 199/8 | 178/9 | 76/17 | 51/24 | 85/16 | **22**/**33** | **18**/**38** | 60/21 | 50/25 | 50/25 | 33/30 | 25/29 |
| *animal* | 44/34 | 42/35 | 25/75 | 29/60 | 19/80 | **11**/**90** | 17/80 | 28/75 | 20/77 | 30/55 | 20/79 | **8**/**92** |
| *skating1* | 99/29 | 98/29 | 57/60 | 80/52 | 150/13 | **9**/**90** | 12/88 | 133/22 | 129/22 | 99/24 | 19/75 | **6**/**93** |
| *singer1* | 51/54 | 24/71 | 51/54 | 17/80 | 13/83 | 8/90 | 13/83 | 22/72 | **5**/**95** | **5**/**95** | 14/83 | 12/83 |
| *skating2* | 56/46 | 57/45 | N/W | 93/19 | 80/22 | **42**/**51** | 43/50 | 68/39 | 97/19 | 74/26 | 99/18 | **28**/**67** |
| *Average* | 63/41 | 56/44 | 53/50 | 49/50 | 59/45 | 20/66 | **19**/**68** | 55/45 | 61/47 | 51/51 | 34/63 | **15**/**71** |

**Analysis of the Parameters:** Our method is less sensitive to the initialization of the IBB and OBB, as demonstrated in Fig.5(a). To get the standard deviation, we changed the initial center positions of the IBB and OBB by adding random noises, to a maximum of 5 pixels. Then, we obtained 10 center location errors from 10 different initialization settings.

We also tested the parameter sensitivity of $\alpha$ and $\beta$ in (11). For this experiment, we obtained the standard deviation of center location errors of 10 tests. To make 10 tests, we added $\alpha$ and $\beta$ with 10% noise. As shown in Fig.5(b), our method is not much sensitive to the parameter settings.

### 5.2 Comparison with Other Methods

The proposed method (DBB) was compared with 11 different state-of-the-art tracking methods [27,37]: ABCShift [36], MC [19], IVT [30], MIL [3], BHMC [22], BHT [26], HT [12], LGT [6], VTD [20], VTS [21], and TLD [18], where VTD and VTS are state-of-the-art trackers that use *color* information and BHMC, BHT, HT, and LGT are trackers that are designed especially for highly *non-rigid* targets. We tested 15 sequences[1].

[1] 12 sequences are publicly available. Only 3 sequences (*basketball*, *lazysong*, and *fx*) were made by us.

**Table 3. Quantitative comparison of tracking results** with other methods. In this experiment, other tracking methods utilize the **OBB representation**.

| | ABCShift | BHMC | BHT | MIL | IVT | VTD | VTS | MC | HT | LGT | TLD | DBB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *basketball* | 99/28 | 63/**40** | 157/20 | 101/29 | 237/11 | 177/15 | **62**/**40** | 133/22 | 169/21 | 170/16 | 116/26 | **36**/**65** |
| *lazysong* | 60/45 | 74/43 | 115/27 | 48/53 | 43/55 | **24**/**65** | 29/61 | 94/31 | 137/22 | 71/45 | 34/57 | **17**/**69** |
| *fx* | 63/42 | 29/65 | N/W | 75/36 | **27**/**69** | 39/59 | 28/68 | 37/61 | 70/41 | 70/41 | 39/59 | **25**/**70** |
| *diving* | 26/31 | **18**/**45** | 74/25 | 88/19 | 91/16 | 70/27 | 98/14 | 29/39 | 83/20 | 29/49 | 84/20 | **16**/**46** |
| *gymnastics* | 20/64 | **7**/**87** | N/W | 22/60 | 27/51 | **10**/**80** | 92/35 | 76/40 | 102/25 | 98/31 | 74/41 | 16/68 |
| *faceocc* | 29/70 | 30/69 | 29/71 | 25/75 | 27/74 | 7/88 | **6**/**89** | 31/69 | 13/81 | 19/78 | 19/78 | **6**/**88** |
| *twinings* | 20/74 | 29/69 | 43/52 | 10/85 | 32/64 | **7**/**91** | **7**/90 | 27/70 | 36/60 | 24/73 | 17/79 | **6**/**89** |
| *tiger1* | 85/30 | 34/39 | 21/58 | 16/60 | 90/29 | 17/59 | 15/60 | 32/41 | 40/39 | 16/60 | **10**/**69** | **13**/**65** |
| *david* | 38/37 | 30/34 | 15/50 | 29/47 | 13/58 | 11/60 | 10/60 | 39/33 | 9/62 | **5**/**72** | 7/64 | **6**/**64** |
| *shaking* | 38/79 | 50/60 | 20/85 | 37/77 | 130/20 | 7/87 | **6**/**93** | 100/21 | 12/79 | 12/79 | **5**/**94** | **6**/90 |
| *soccer* | 71/13 | 150/11 | 43/24 | 43/24 | 104/16 | **22**/**33** | **17**/**39** | 49/26 | 55/22 | 49/26 | 33/30 | 25/29 |
| *animal* | 30/40 | 59/30 | 29/47 | 29/47 | 22/51 | **13**/**88** | 15/86 | 30/46 | 25/49 | 35/49 | 17/81 | **8**/**92** |
| *skating1* | 130/20 | 129/20 | 77/54 | 80/52 | 150/13 | **11**/**88** | 12/**88** | 172/19 | 126/21 | 99/22 | 22/70 | **6**/**93** |
| *singer1* | 22/78 | 55/52 | 48/61 | 24/77 | **3**/**99** | 5/96 | 5/96 | 84/39 | 51/54 | **4**/**98** | 53/52 | 12/83 |
| *skating2* | 80/31 | 52/45 | 49/47 | 95/29 | 157/19 | 30/64 | **29**/**66** | 36/59 | 174/12 | 160/18 | 85/31 | **28**/**67** |
| *Average* | 54/45 | 53/47 | 55/47 | 48/51 | 76/43 | 30/**66** | **28**/65 | 64/41 | 73/40 | 57/50 | 41/56 | **15**/**71** |

For sufficient comparison, we used 2 different settings of the bounding box like Tables 2 and 3. Tables 2 and 3 show the quantitative evaluation of the tracking results. Our method always used the DBB representation whereas other methods used the inner and outer bounding box representations to produce results in Tables 2 and 3, respectively. These tracking results indicate that our method is robust to track deformable target objects, as the drifting problem is effectively resolved using the DBB representation. In terms of the center location error, our method outperformed the recent state-of-the-art tracking methods especially for highly non-rigid objects, which are BHMC, BHT, HT, and LGT. Our method was also better than the other tracking methods in terms of the sucess rate, because our method successfully tracked the targets to the last frame although it produces slightly inaccurate OBB. On the other hand, other tracking methods fail to track the target and drift into the background, which make the low success rate. The tracking results demonstrate that a single bounding box representation is not adequate to represent highly non-rigid objects. In addition, our method produced better results than color-based tracking methods (BHMC, VTD, VTS, and MC) because our method did not consider the ambiguous regions while calculating color histograms. Our method only calculates color histograms for the region inside the IBB and the region outside of the OBB. Notably, conventional tracking methods using a single bounding box yielded very different tracking results depending on the representation type of a single bounding box (i.e., IBB representation in Table 2 and OBB representation in Table 3). Conversely, our method does not depend on the representation type of a single bounding box because it uses both the IBB and OBB.

Fig.7(e) shows the tracking results in the *gymnastics* seq. In the sequence, VTD showed the best result, but its distance from the target became wider when the gymnast turned and changed her pose fast. Fig.6 shows the results

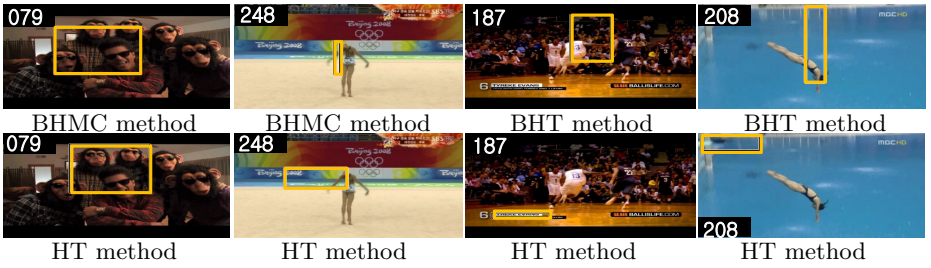



**Fig. 6.** Tracking results of the methods especially for highly non-rigid objects


**Fig. 7. Qualitative comparison of the tracking results with other state-of-the-art tracking methods.** The red and blue boxes give the results of the proposed method (Combination of the IBB and OBB). The yellow, white, orange, green, and pink boxes give the results of MCMC method using the IBB representation, VTD, VTS, IVT, and MIL using the OBB representation, respectively.

of BHMC, BHT, and HT, which are designed especially for highly non-rigid objects, and demonstrates that these methods also frequently failed to track the targets when there was severe deformation of the targets.

In Fig.7(a), the *lazysong* seq., which includes some objects similar in appearance to the target object, is tested. In the case of other methods, their bounding boxes expanded and included some background objects. However, our method showed the most accurate tracking performance among these methods. Fig.7(b) shows the tracking results of the *fx* seq. The target person was severely occluded by another person who wore clothes of the same color as that of the target. Whereas some trajectories were hijacked by the other person, our method successfully tracked the target. Fig.7(c) shows the tracking results of the *diving* seq. When the woman started spinning, our method continued to track the woman while the other methods failed to track it. Fig.7(d) shows the tracking results in the *basketball* seq. Our method maintained the trajectory of the target. However, the other methods experienced drifting problems, as they had a larger background part than the target part in their bounding boxes in frame #187 and frame #636.

## 6 Conclusion and Discussion

In this paper, we propose a new bounding box representation called the double bounding box representation. The proposed bounding box represents the target as the combination of the inner and outer bounding boxes and does not need to deal with the ambiguous regions, which include both the target and the background, at the same time. Hence, the method greatly improves tracking accuracy without additional computational cost.

IBB and OBB can be exactly same for a rigid object. Hence equations of the belief and plausibility in (1) and (8) include the equality. In this case, our tracking performance is similar to the single BB based tracking methods.

Although the color feature is applied to our method, other features can be also used in the method. For example, histogram of gradient (HOG) can be used in the faceocc sequence, where a region inside the IBB is represented by HOG of eyes, nose, and mouse, but a region outside the OBB by HOG of the book.

The basic idea behind our method is intuitive and can be implemented without any theories. However, there are few works that try to find a theoretical basis of the idea. As the theoretical basis, in this paper, we present the DS theory and prove that our approach is optimal in terms of the DS theory.

**Acknowledgements.** This work was partly supported by the ICT R&D program of MSIP/IITP, Korea [14-824-09-006, Novel Computer Vision and Machine Learning Technology with the Ability to Predict and Forecast] and the European Research Council (ERC) under the project VarCity (#273940). The authors gratefully acknowledge support by Toyota.

## References


1. Adam, A., Rivlin, E., Shimshoni, I.: Robust fragments-based tracking using the integral histogram. In: CVPR (2006)
2. Avidan, S.: Ensemble tracking. PAMI 29(2), 261–271 (2007)
3. Babenko, B., Yang, M., Belongie, S.: Visual tracking with online multiple instance learning. In: CVPR (2009)
4. Bao, C., Wu, Y., Ling, H., Ji, H.: Real time robust l1 tracker using accelerated proximal gradient approach. In: CVPR (2012)
5. Birchfield, S.: Elliptical head tracking using intensity gradients and color histograms. In: CVPR (1998)
6. Cehovin, L., Kristan, M., Leonardis, A.: An adaptive coupled-layer visual model for robust visual tracking. In: ICCV (2011)
7. Collins, R.T., Liu, Y., Leordeanu, M.: Online selection of discriminative tracking features. PAMI 27(10), 1631–1643 (2005)
8. Comaniciu, D., Ramesh, V., Meer, P.: Real-time tracking of non-rigid objects using mean shift. In: CVPR (2000)
9. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Statist. 38(2), 325–339 (1967)
10. Faux, F., Luthon, F.: Robust face tracking using colour dempster-shafer fusion and particle filter. In: FUSION (2006)
11. Glenn, S.: A mathematical theory of evidence. Princeton University Press (1976)
12. Godec, M., Roth, P.M., Bischof, H.: Hough-based tracking of non-rigid objects. In: ICCV (2011)
13. Han, B., Davis, L.: On-line density-based appearance modeling for object tracking. In: ICCV (2005)
14. Hare, S., Saffari, A., Torr, P.H.S.: Struck: Structured output tracking with kernels. In: ICCV (2011)
15. Isard, M., Blake, A.: ICONDENSATION: Unifying low-level and high-level tracking in a stochastic framework. In: Burkhardt, H.-J., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1406, pp. 893–908. Springer, Heidelberg (1998)
16. Jepson, A.D., Fleet, D.J., Maraghi, T.F.E.: Robust online appearance models for visual tracking. PAMI 25(10), 1296–1311 (2003)
17. Jia, X., Lu, H., Yang, M.H.: Visual tracking via adaptive structural local sparse appearance model. In: CVPR (2012)
18. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. PAMI 34(7), 1409–1422 (2012)
19. Khan, Z., Balch, T., Dellaert, F.: MCMC-based particle filtering for tracking a variable number of interacting targets. PAMI 27(11), 1805–1918 (2005)
20. Kwon, J., Lee, K.M.: Visual tracking decomposition. In: CVPR (2010)
21. Kwon, J., Lee, K.M.: Tracking by sampling trackers. In: ICCV (2011)
22. Kwon, J., Lee, K.M.: Tracking of a non-rigid object via patch-based dynamic appearance modeling and adaptive basin hopping monte carlo sampling. In: CVPR (2009)
23. Li, X., Dick, A., Shen, C., Zhang, Z., van den Hengel, A., Wang, H.: Visual tracking with spatio-temporal dempstershafer information fusion. TIP (2013)
24. Mei, X., Ling, H.: Robust visual tracking using l1 minimization. In: ICCV (2009)
25. Munoz-Salinas, R., Medina-Carnicer, R., Madrid-Cuevas, F., Carmona-Poyato, A.: Multi-camera people tracking using evidential filters. Ann. Math. Statist. 50, 732–749 (2009)

26. Nejhum, S.M.S., Ho, J., Yang, M.H.: Visual tracking with histograms and articulating blocks. In: CVPR (2008)
27. Pang, Y., Ling, H.: Finding the best from the second bests- inhibiting subjective bias in evaluation of visual tracking algorithms. In: ICCV (2013)
28. Pérez, P., Hue, C., Vermaak, J., Gangnet, M.: Color-based probabilistic tracking. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part I. LNCS, vol. 2350, pp. 661–675. Springer, Heidelberg (2002)
29. Ramanan, D., Forsyth, D., Zisserman, A.: Tracking people by learning their appearance. PAMI 29(1), 65–81 (2007)
30. Ross, D.A., Lim, J., Lin, R., Yang, M.: Incremental learning for robust visual tracking. IJCV 77(1), 125–141 (2008)
31. Santner, J., Leistner, C., Saffari, A., Pock, T., Bischof, H.: Prost: Parallel robust online simple tracking. In: CVPR (2010)
32. Sevilla-Lara, L., Learned-Miller, E.: Distribution fields for tracking. In: CVPR (2012)
33. Smith, K., Gatica-Perez, D., Odobez, J.M.: Using particles to track varying numbers of interacting people. In: CVPR (2005)
34. Stalder, S., Grabner, H., Van Gool, L.: Cascaded confidence filtering for improved tracking-by-detection. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part I. LNCS, vol. 6311, pp. 369–382. Springer, Heidelberg (2010)
35. Stenger, B., Woodley, T., Cipolla, R.: Learning to track with multiple observers. In: CVPR (2009)
36. Stolkin, R., Florescu, I., Baron, M., Harrier, C., Kocherov, B.: Efficient visual servoing with the abcshift tracking algorithm. In: ICRA (2008)
37. Wu, Y., Lim, J., Yang, M.H.: Online object tracking: A benchmark. In: CVPR (2013)
38. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Comput. Surv. 38(4) (2006)