

Robust Target Model Update for Mean-shift Tracking with Background Weighted Histogram

Yong-Hyun Jang¹, Jung-Keun Suh^{2,+}, Ku-Jin Kim³ and Yoo-Joo Choi^{2,+*}

¹Department of Stereoscopic Media, Korean German Institute of Technology,
Deugnchon-dong, Gangseo-gu, Seoul, South Korea
[e-mail: yhjang0522@gmail.com]

²Department of Newmedia Content, Seoul Media Institute of Technology,
⁺Immersive Media Lab., Seoul Media Institute of Technology,
Deugnchon-dong, Gangseo-gu, Seoul, South Korea
[e-mail: jksuh@smit.ac.kr, yjchoi@smit.ac.kr]

³School of Computer Science & Engineering, Kyungpook National University
80 Daehak-ro, Buk-gu, Daegu, South Korea
[e-mail: kujinkim@knu.ac.kr]

* Corresponding author: Yoo-Joo Choi

*Received October 20, 2015; revised December 31, 2015; accepted February 3, 2016;
published March 31, 2016*

Abstract

This paper presents a target model update scheme for the mean-shift tracking with background weighted histogram. In the scheme, the target candidate histogram is corrected by considering the back-projection weight of each pixel in the kernel after the best target candidate in the current frame image is chosen. In each frame, the target model is updated by the weighted average of the current target model and the corrected target candidate. We compared our target model update scheme with the previous ones by applying several test sequences. The experimental results showed that the object tracking accuracy was greatly improved by using the proposed scheme.

Keywords: visual object tracking, mean-shift tracking, target model update

This research was supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Industry Optimized Human Resource Development Program 2014 and by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2013R1A1A2A10004391). A preliminary version of this paper was presented at APIC-IST 2015, and was selected as an outstanding paper.

¹ Current affiliation : Business IT Section, Lotte Data Communication Co., Seoul, Korea

1. Introduction

Image-based object tracking technologies have been widely used in various application fields such as broadcasting, intelligent surveillance system, traffic control system, military security system, etc. Among various approaches, the kernel-based mean-shift tracking[1] has attracted many attentions as an effective target object tracking method, since it requires lower computational cost compared to other tracking approaches such as particle filtering and optical flow.

In the kernel-based mean-shift tracking, the target model is defined in the first frame and the target candidate is estimated in the subsequent frame. The target model means the probability density function (pdf) to characterize the target object such as the color pdf. In the subsequent frame, the target candidates are represented by the color pdf that depicts the target candidate regions. To achieve the low-computational cost, the discrete densities, i.e., m -bin histograms are frequently used. Furthermore, a similarity function between the target model and the target candidate is defined. The local maxima of the similarity function in the image indicate the presence of objects in the subsequent frame that looks similar to the target model defined in the first frame. To track a target object in the subsequent frame, we find the best target candidate by using the local maxima of the similarity function in each frame.

A large amount of researches has been conducted to increase accuracy and performance of tracking in the sequence of images of real-life situation. Those images include various obstacles, such as the cluttered background, illumination changes, occlusions, etc, that cause the degradation of the tracking accuracy. To improve the tracking accuracy, various approaches in the target model construction[2-3] have been proposed, where they reduce impacts of the background clutters by considering the background colors. Target model update schemes[4-5] have been attempted to reflect changes of the target object colors due to the illumination changes. Especially, in the outdoor environment, the importance of the target model update has been increased due to widely changeable illumination conditions.

Wu et al.[4] presented the model update method based on the probabilistic covariance tracking. Li et al.[5] proposed the model update method in the kernel-based mean-shift tracking without background-weighted histogram. In their approach, the target model is updated in every frame by weighted averaging the target model and the current target candidate. Approches using the multi-mode kernel or hybrid approaches combined with particle filtering have been proposed for the robust object tracking even in the occlusion circumstance[6-7]. Tracking multiple objects[8] has been also researched as a major issue of the image-based object tracking in the last decade.

This paper focuses on designing the effective target model update method, which can be used in kernel-based mean-shift tracking with background weighted histograms [2-3] to robustly track a target object in the various real-life situations. In the case of applying background weighted histograms for constructing the initial target model[1-3], the target model represents the color pdf having high probabilities to unique colors of the target object. On the other hand, the target candidate is the color pdf with high probabilities not only to unique colors of the target object but also to the background colors within the kernel. Therefore, Li et al.[5]'s target model update method which has been widely used in the kernel-based object tracking can not be applied to kernel-based tracking with background weighted histograms. It is necessary to extract the unique colors of the target object from the target candidate in each frame image in order to utilize the target candidate for the target model

update. In this paper, we propose a novel target model update scheme which uses the corrected target candidates based on the back-projection weight for the kernel-based tracking with background weighted histograms.

The remains of the paper are organized as follows. Section 2 introduces the mean-shift object tracking with background-weighted histogram that has shown the robust tracking results. The proposed model update scheme for the kernel-based mean-shift tracking with background-weighted histogram is addressed in section 3. Experiment results using the proposed method are given in section 4 and we conclude this paper in section 5.

2. Mean-shift object tracking with background weighted histogram

2.1 Target model construction with background weighted histogram

The kernel-based mean-shift tracking is used to track an ellipsoidal region of a target object based on a kernel weighted histogram of colors[2]. In an initial frame, a target model $q = \{q_u\}_{u=1..m}$ with background weights is defined by the probability of the feature $u=1..m$ computed as

$$q_u = C v_u \sum_{i=1}^n k(\|x_i\|^2) \delta[b(x_i) - u], \quad (1)$$

where $\{x_i\}_{i=1..n}$ mean the normalized pixel locations in the region defined as the target model and $b(x_i)$ is the index of the histogram bin corresponding to the color of the pixel x_i . $k(\|x_i\|)$ is a convex and monotonic decreasing kernel profile which assigns a smaller weight to locations that are farther from the center of the target. δ is the Kronecker delta function and Constant C is expressed as

$$C = \frac{1}{\sum_{i=1}^n k(\|x_i\|^2) \sum_{u=1}^m v_u \delta[b(x_i) - u]}. \quad (2)$$

v_u is a background weight for reducing background interference. v_u is computed using the background histogram $\{o_u\}_{u=1..m}$ (with $\sum_{u=1}^m o_u = 1$) and its smallest nonzero entry o^* . The background weights are computed in a region around the target as shown in Fig. 1. We assumed that the background region has three times bigger area than that of the target object. The weights v_u is defined by

$$\{v_u = \min(\frac{o^*}{o_u}, 1)\}_{u=1..m}. \quad (3)$$



Fig. 1. The background region around the target

In the target model, the probabilities of major background colors are suppressed while those of the unique colors of the target object are enhanced by using background weight v_u . **Fig. 2** shows the target models - color histograms for the kernel region before and after applying background weight v_u . It represents that the colors of the target object are emphasized after applying background weight v_u .

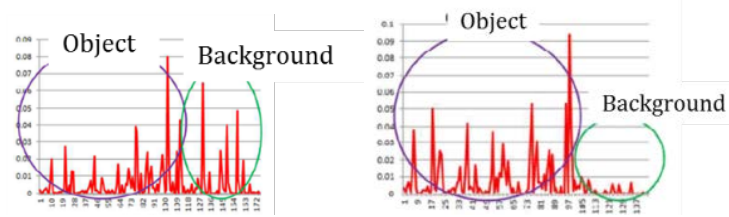


Fig. 2. Target model before (left) and after (right) applying background weight v_u

2.2 Target candidate representation

The target is tracked by comparing the similarity between the target model q and a target candidate $p = \{p_u\}_{u=1..m}$ in a sequence of video frames. Let $\{x_i\}_{i=1..n^h}$ be the pixel locations of the target candidate, centered at y in the current frame. Using the same kernel profile $k(x)$, but with bandwidth h , the probability of the feature $u=1..m$ in the target candidate is given by

$$p_u(y) = C_h \sum_{i=1}^{n^h} k\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \delta[b(x_i) - u], \quad (4)$$

where

$$C_h = \frac{1}{\sum_{i=1}^{n^h} k\left(\left\|\frac{y-x_i}{h}\right\|^2\right)} \quad (5)$$

is the normalization constant.

2.3 Mean-shift algorithm

The search for the new target location y_1 in the current frame starts at the estimated location y_0 of the target in the previous frame and is repeatedly computed as Eq. (6).

$$y_1 = \frac{\sum_{i=1}^{n^h} x_i w_i g\left(\left\|\frac{y_0-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n^h} w_i g\left(\left\|\frac{y_0-x_i}{h}\right\|^2\right)}, \quad (6)$$

where $g(x)$ is the shadow of the kernel profile $k(x)$, i.e. $g(x) = -k'(x)$, and w_i is the back-projection weight given as

$$w_i = \sum_{u=1}^m \delta[b(x_i) - u] \sqrt{\frac{q_u}{p_u(y)}}. \quad (7)$$

If $\|y_1 - y_0\| < \varepsilon$, y_l is defined as the center of the most similar region to the target object in the new frame. Otherwise, y_l is assigned to y_0 and y_l is recomputed using Eq. (6).

3. Target model update based on back-projection weights

3.1 Overall tracking procedure

The visual object tracking procedure consists of four processing steps as shown in Fig. 3. In the initialization step, the target model is constructed as explained in section 2.1. In the second step, the best target candidate is found by repeated computations of Eq. (4) and Eq. (6) in each video frame. The changed size of the target object is estimated in the third step and the target model update is finally conducted based on the best target candidate region which is determined by step 2 and step 3. For the scale estimation (step 3), we applied Jeyaker's scale estimation method[3] and this paper focuses on how to robustly update the target model in order to enhance the tracking correctness. The proposed target model update method is based on target candidate correction using the back-projection weight. The algorithm details are presented in the next section.

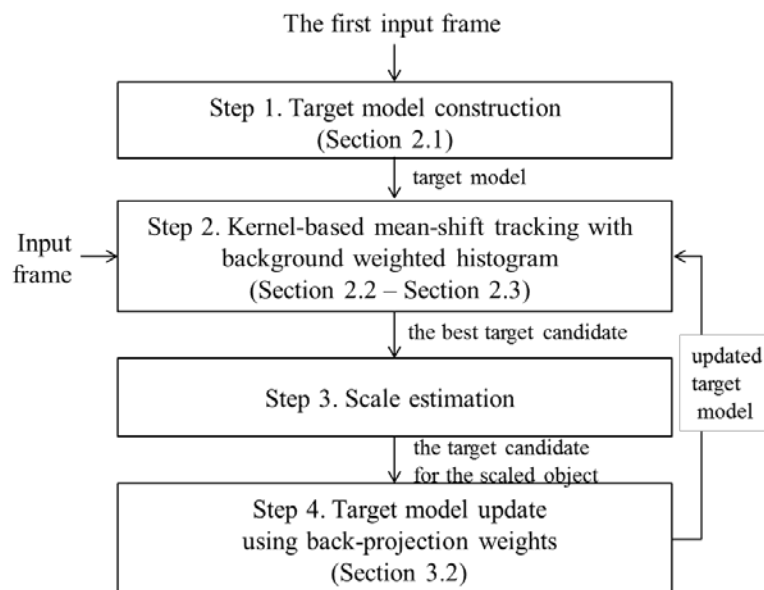


Fig. 3. Overview of the tracking procedure

3.2 Target candidate correction and target model update

The target object is tracked by comparing the similarity between the target model q and a target candidate p in a sequence of video frames. After the final location of the kernel for the target candidate p^* is determined in each frame, the target candidate p^* is reconstructed by selecting the kernel pixels that have larger back-projection weight w_i than a threshold ε . The back-projection weight w_i for each pixel x_i in the kernel is defined as Eq. (7). Fig. 4 shows a sample back-projection image, where white pixels having large weight w_i . As shown in Fig. 4, the pixels having large back-projection weight w_i are included in the target object region, which can be used to separate a target object from the background.

The proposed target model update method was basically devised based on the fact that the target candidates in each frame have high probabilities to both colors of target object and the background colors within the kernel. Therefore, it is necessary to attenuate the probabilities to background colors in the target candidate in order to update the target model by weighted average between the target model and the target candidate. In the proposed scheme, we select the pixels that have larger back-projection weight w_i than a threshold ϵ is to exclude the background pixels in the kernel.

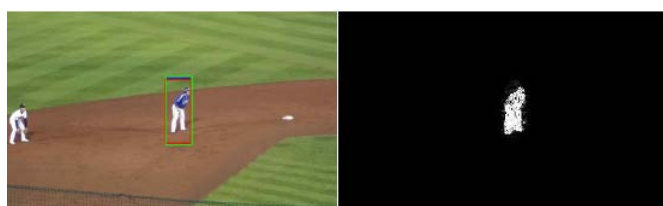


Fig. 4. The kernel (left) and back-projection image (right) for the 298th frame of the baseball 2 test sequence.

Algorithm 1 presents the procedure for the correction of the target candidate. Each pixel in the target candidate is checked whether its colors corresponds to the major object color in the target model q_u and its weight w_i is greater than the threshold. The only pixels that are satisfied with both two conditions are used in correcting the target candidate (Refer line 10). In line 14, the corrected target candidate p^* is normalized.

Algorithm 1 : Correction of target candidate based on back-projection weights

Input : target model q ,

The best target candidate p in the frame

target candidate region

threshold ϵ, \mathfrak{D}

Output : Corrected target candidate p^*

1. Set all $p_u^*(y_0)$ to 0
2. for (all pixels in target candidate region) {
3. if ($q_u > \mathfrak{D}$) {
4. $w_i = \sum_{u=1}^m \sqrt{\frac{q_u}{p_u(y_0)}} \delta[b(x_i) - u];$
5. }
6. else
7. $w_i = 0;$
8. }
9. if ($w_i > \epsilon$) {
10. $p_u^*(y_0) = p_u^*(y_0) + k \left(\left\| \frac{y_0 - x_i}{h} \right\|^2 \right) \delta[b(x_i) - u];$
11. }
12. }
- 13.
14. $p^* = \text{normalize}(p^*), p^* = \{p_u^*\}_{u=1..m}$

Finally, the target model q is updated by weighted average with the corrected target candidate p^* . That is, the updated target model q_u^* is computed by Eq. 8.

$$q_u^* = (1 - \tau)q_u + \tau p_u^*, \quad u = 1..m \quad (8)$$

where the weight τ is determined heuristically.

4. Experimental Results

In the experiments, we compared the tracking errors of our target model update scheme with those of the previous scheme using five test video sequences. **Fig. 5** shows five test sequences. Four video sequences (Baseball1, Baseball2, Cam24, Xeron2) were produced by authors using Sony CX560 camcorder and AXIS PTZ camera. Corridor test sequences are publicly available from CAVIAR project site¹. Each test sequence has different features which could affect tracking performance. In the case of Baseball1 test sequence, colors of the target object (player) and the background are similar. Baseball2 includes a fast moving target object and Cam24 sequence has a complex background environment. In Xeron2 and Corridor test sequences, the size of the target object (a walking person) is largely changed.

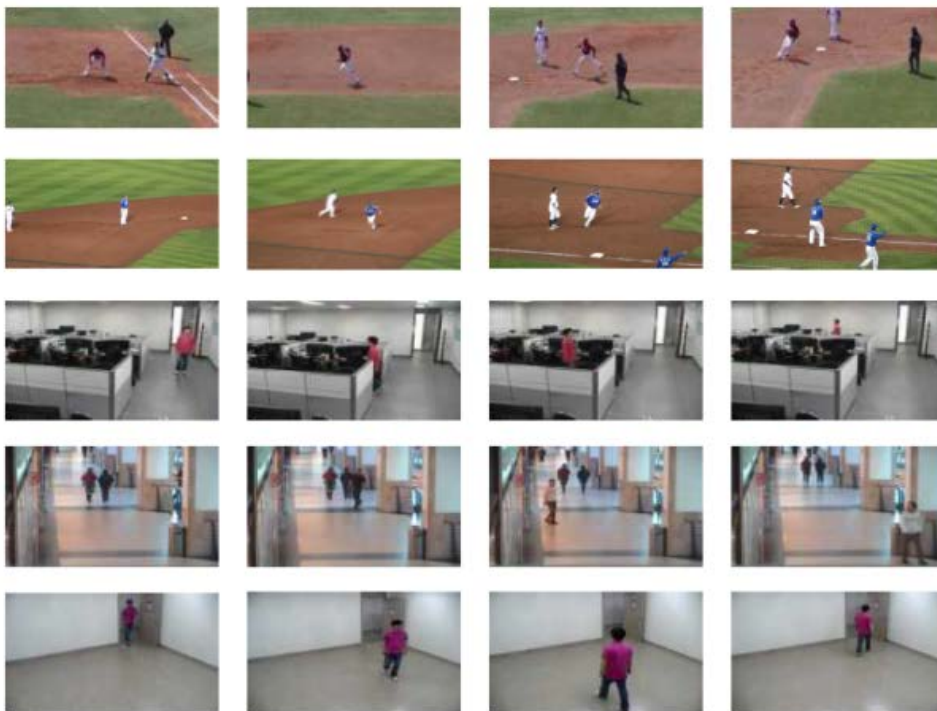


Fig. 5. Test video sequences. From top to bottom row: Baseball1, Baseball2, Cam24, Corridor and Xeron2 sequences

¹ <http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/>

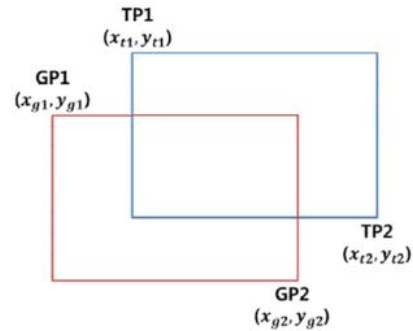
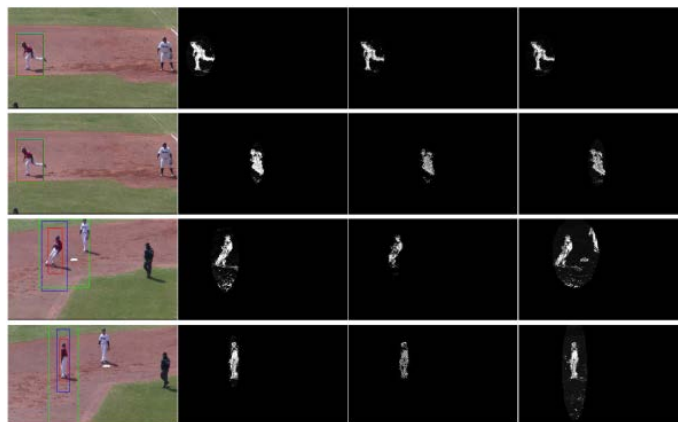


Fig. 6. Bounding box (in red) of a target object extracted from the ground truth data and that (in blue) determined by the tracking result.

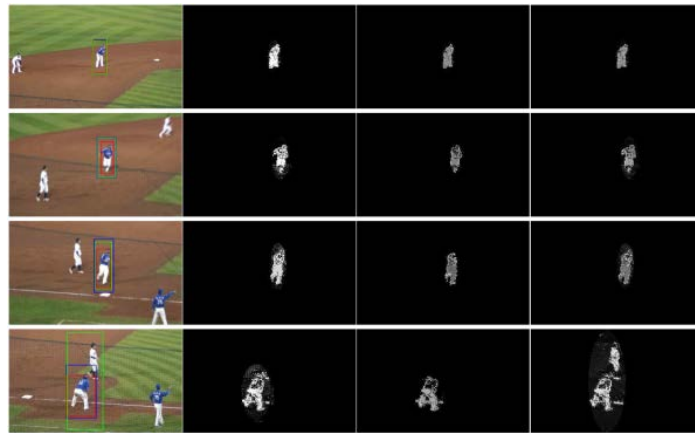
In order to measure tracking errors, we segmented a target object region manually and extracted a tight bounding box for the target object in each video frame. **Fig. 6** shows two bounding boxes – The red one is a bounding box of a target object extracted from the ground truth data. The blue one is a tracking result kernel. Let $GP1(x_{g1}, y_{g1})$, $GP2(x_{g2}, y_{g2})$, $TP1(x_{t1}, y_{t1})$ and $TP2(x_{t2}, y_{t2})$ be the left-top, right-bottom points of two bounding boxes as shown in **Fig. 6**, respectively. The tracking error E in an image frame is defined by Eq. (9).

$$E = |x_{g1} - x_{t1}| + |y_{g1} - y_{t1}| + |x_{g2} - x_{t2}| + |y_{g2} - y_{t2}| \quad (9)$$

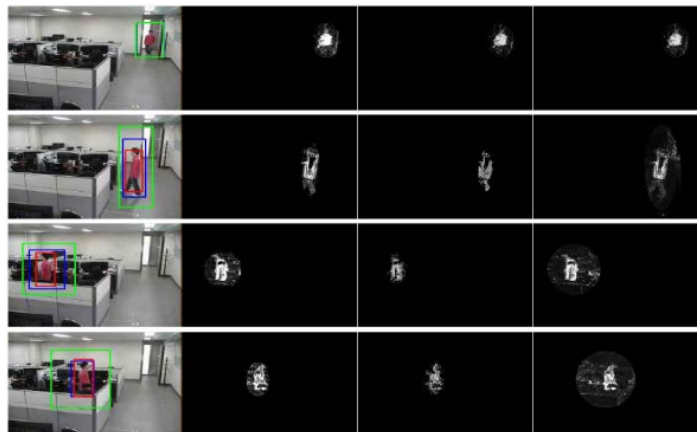
Fig. 7 compares the tracking results by showing the tracking kernels and back-projection images after conducting tracking algorithms without target model update, with ours and with Li's scheme, respectively. In tracking result kernel image (the leftmost) of each row, blue, red, green boxes represent the tracking results without target model update, with ours and with Li's method, respectively. As shown in **Fig. 7**, the tracking method with our target model update method tracked the target more correctly and robustly. In back-projection images, brighter pixels represent the higher values of back-projection weight. After applying our method, white pixels on the background region were more clearly removed. These changes help our method successfully track the given target object until the end of the test sequence.



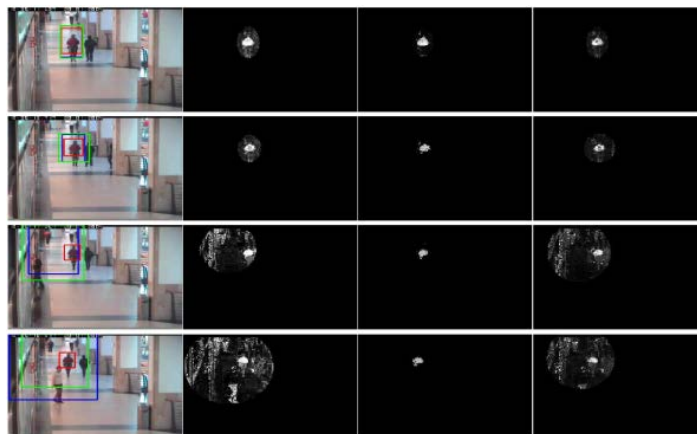
(a) Baseball1



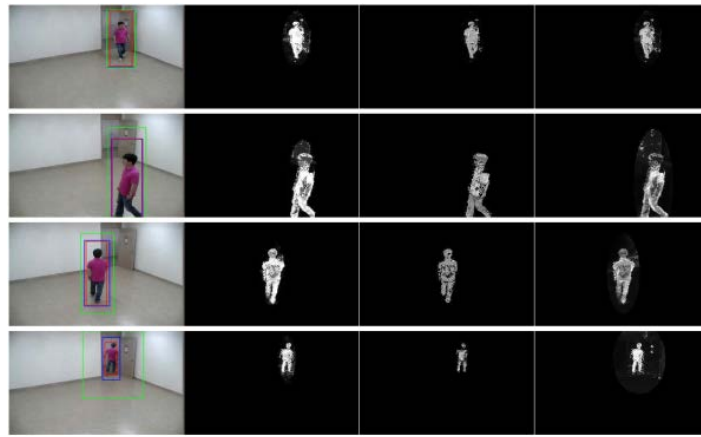
(b) Baseball2



(c) Cam24



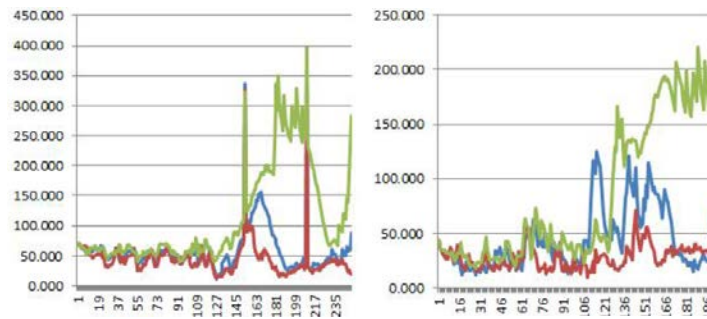
(d) Corridor



(e) Xeron2

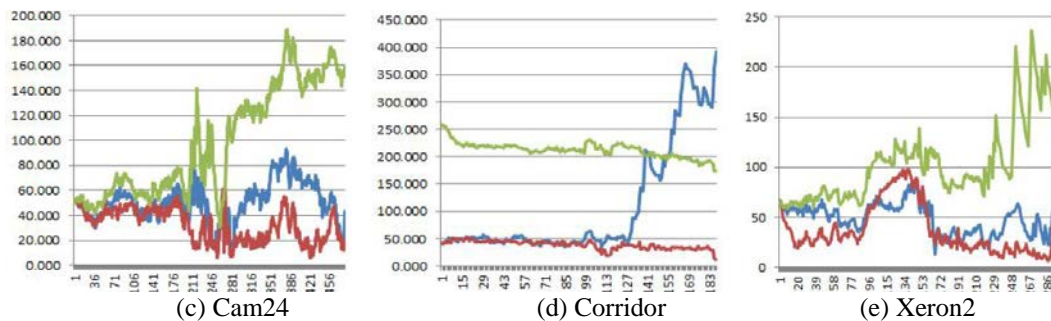
Fig. 7. Tracking result kernels and back-projection images after applying tracking methods without target model update, with ours and with Li's scheme, respectively (From left to right column).

Fig. 8 depicts tracking errors in all video frames of five test sequences. Blue, red and green lines represent tracking errors of methods without target model update, with our target model update, and with Li's, respectively. **Fig. 8** shows that our method significantly improves the tracking accuracy and robustness for every test sequence.



(a) Baseball1

(b) Baseball2



(c) Cam24

(d) Corridor

(e) Xeron2

Fig. 8. Tracking errors without target model update (blue), with Li's target model update scheme (green), with ours (red).

Table 1 compares the tracking errors of test sequences, where each number represents the tracking error averaged over all the tested frames of each video. In **Table 1**, the accuracy improvement (AI) ratio was computed as follows:

$$AI = \frac{E_{no} - E_{ours}}{E_{ours}} \times 100, \quad (10)$$

where E_{no} and E_{ours} are the average tracking errors before and after applying our model update scheme, respectively. The proposed target model update increased the tracking accuracy with 72.45% in average.

Table 1. Comparison of tracking errors before and after applying the target model update

Test seq.	Without target model update	Applying Li's target model update	Applying proposed target model update	accuracy improvement ratio of proposed method (%)
Baseball1	61.82	88.82	46.93	31.71
Baseball2	44.42	76.63	27.49	61.57
Cam24	51.01	98.65	32.32	57.82
Corridor	108.52	212.72	38.71	180.35
Xeron2	43.38	102.30	36.99	30.78

5. Conclusions

Kernel-based mean-shift tracking with background weighted histogram is known to be stable to various tracking obstacles such as background clutters and illumination changes. However, tracking is still failed easily in the cases of rapid illumination changes and the cluttered background, which happen frequently in the outdoor environment. In order to overcome those problems, the accuracy and robustness of the target model update scheme is required. Therefore, in this paper, a novel target model update method was proposed for the kernel-based mean-shift tracking with background weighted histogram.

Using the comparative experiments, we validated that the proposed target model update enhanced the tracking accuracy and robustness. For five test sequences, the experiments showed that the accuracy of the proposed method was improved by 72.45% in average.

For more accurate tracking, the efficient and robust scaling schemes are also needed. Therefore, we are going to analyze previous scaling estimation schemes and propose an efficient and robust scaling scheme as a future work.

References

- [1] D. Comaniciu, V. Ramesh, P. Meer, “Kernel-Based Object Tracking,” *IEEE Trans. of Pattern Analysis and Machine Intelligence*, vol 25, no. 5, pp.563-577, 2003. [Article \(CrossRef Link\)](#)
- [2] J. Ning, L. Zhang, D. Zhang, C. Wu, “Robust mean-shift tracking with corrected background-weighted histogram,” *IET Comput. Vis.*, vol. 6, no. 1, pp. 62-69, 2012. [Article \(CrossRef Link\)](#)
- [3] J. Jeyakar, R. V. Babu, K. R. Ramakrichna, “Robust object tracking with background-weighted local kernels,” *Computer Vision and Image Understanding*, vol. 112, no. 3, pp. 296-309, December 2008. [Article \(CrossRef Link\)](#)
- [4] Y.Wu, J. Cheng, J. Wang, H. Lu, et al., “Real-time Probabilistic Covariance Tracking with Efficient Model Update,” *IEEE Trans. Image Process.* vol. 21,no. 5, pp.2824-2837, 2012. [Article \(CrossRef Link\)](#)
- [5] L. Li, Z. Feng, “An efficient object tracking method based on adaptive nonparametric approach,” *Opto-electronics Review*, vol. 13, no. 4, pp. 325-330, 2005. [Article \(CrossRef Link\)](#)
- [6] Z. Khan, I. Y.-H. Gu, and A. Backhouse, “Joint particle filters and multi-mode anisotropic mean shift for robust tracking of video objects with partitioned areas,” in *Proc. of IEEE Int. Conf. Image Process*, pp. 4077–4080, November 2009. [Article \(CrossRef Link\)](#)
- [7] Z. Khan, I. Y.H.Gu, An G. Backhouse, “Robust Visual Object Tracking Using MultiMode Anisotropic Mean Shift and Particle Filters,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol.21, no. 1, pp. 74-87, January 2011. [Article \(CrossRef Link\)](#)
- [8] H. Fang, J.W. Kim, and J.W. Jang, “A Fast Snake Algorithm for Tracking Multiple Objects,” *Journal of Information Processing Systems*, vol. 7, no. 3, pp. 519-530, September 2011. [Article \(CrossRef Link\)](#)



Yonghyun Jang received the M.S degrees in Stereoscopic Media Technology from Korea and German Institute of Technology. Seoul, Korea, in 2013 and 2015. His research focuses on Visual object tracking, Machine Learning, Pattern Recognition, etc.



Jung-Keun Suh is a professor in Department of Newmedia at Seoul Media Institute of Technology (SMIT). He gained Bachelor and Master degrees from Plant Sciences, Seoul National University at 1987 and 1989. He gained his doctorate from Department of Biochemistry, University of Texas at Austin (UT-Austin), USA, at 1996, studied for structural biology and proteomics. Before joining LG Life Sciences (LGLS), he was a Post-Doctoral Fellow in Department of Biochemistry at UT-Austin for 3 years. He joined LGLS as a leader of Proteomics Program and then as the Head of Protein Characterization Program from 2000 to 2006. He left LGLS at Nov, 2006, joined and served the director of Biotech Institute of the Korean German Institute of Technology. After joining SMIT, his research focuses on human sensing, bio-data visualization, new media for healthcare, etc.



Ku-Jin Kim is a Professor in the School of Computer Science & Engineering at Kyungpook National University, Korea. Her research interests include computer graphics, parallel programming, computational biology, and geometric/surface modeling. Prof. Kim received a B.S. degree from Ewha Womans University in 1990, an M.S. degree from KAIST in 1992, and a Ph.D. degree from POSTECH in 1998, all in Computer Science. She was a PostDoctoral fellow at Purdue University, USA, in 1998-2000. Prof. Kim also held faculty positions at Ajou University, Korea, and at University of Missouri, St. Louis, USA



Yoo-Joo Choi is a Professor in Department of Newmedia at Seoul Media Institute of Technology(SMIT), Korea. She received her M.S. and Ph.D. degrees in Computer Science from Ewha Womans University in 1991 and 2005, respectively. She was a researcher at R&D Department of KCI Co. and POSDATA Co. in Korea between 1991 and 1999. Her research interests include image processing, computer vision, computer graphics, augmented reality and human-computer interaction.