멀티모드 커널 가중치 기반 객체 추적

김은섭[1○] 김용구[2] 최유주[2*]

[1]전자부품연구원

[2]한독미디어대학원대학교, 뉴미디어콘텐츠학과

(maycos@naver.com, ygkim@kgit.ac.kr, yjchoi@kgit.ac.kr)

Multi-mode Kernel Weight-based Object Tracking

Eun-Sub Kim[1○] Yong-Goo Kim[2] Yoo-Joo Choi[2*]

1 Korea Electronics Technology Institute
2 Department of Newmedia, Korean German Institute of Technology

Abstract

As the needs of real-time visual object tracking are increasing in various kinds of application fields such as surveillance, entertainment, etc., kernel-based mean-shift tracking has received more interests. One of major issues in kernel-based mean-shift tracking is to be robust under partial or full occlusion status. This paper presents a real-time mean-shift tracking which is robust in partial occlusion by applying multi-mode local kernel weight. In the proposed method, a kernel is divided into multiple sub-kernels and each sub-kernel has a kernel weight to be determined according to the location of the sub-kernel. The experimental results show that the proposed method is more stable than the previous methods with multi-mode kernels in partial occlusion circumstance.

요 약

최근, 감시시스템, 게임, 영화등 다양한 분야에서 영상을 이용한 실시간 객체 추적의 필요성이 높아짐에 따라, 커널기반 mean-shift 추적 기법에 대한 관심이 높아지고 있다. 커널 기반 mean-shift 객체 추적에 있어서 주요한 몇 가지 문제점들 중, 첫번째로 추적 목표 객체에 대한 부분 가림 혹은 전체 가림 상황에서의 객체 추적의 문제를 들 수 있다. 본 논문에서는 멀티모드 지역적 커널 가중치를 적용함드로써 부분 가림 상황에서도 안정적드로 객체를 추적할 수 있는 실시간 mean-shift 추적 기법을 제안한다. 제안기법에서는 단일 커널을 사용하는 대신 여러 개의 서브 커널들로 구성된 커널을 사용하고, 각 서브 커널의 위치에 따른 지역적 커널 가중치를 적용한다. 기존의 멀티모드 커널 기반의 방법과 비교한 실험을 통하여 본 제안 방법이 보다 안정적드로 객체를 추적할 수 있음을 보였다.

키워드; 객체 추적, mean-shift 추적, 커널 기반 추적, 가림 처리

Keywords: object tracking; mean-shift tracking; kernel-based tracking, occlusion handling

*corresponding author: Yoo-Joo Choi/Korean German Institute of Technology (yjchoi@kgit.ac.kr)

Received : 2015.07.25 / Review completed : 2nd 2015.08.29 / Accepted : 2015.08.31
DOI : 10.15701/kcgs.2015.21.4.11
ISSN : 1975-7883(Print)/2383-529X(Online)

1. Introduction

Visual object tracking has gained much attention in recent years as demands of intelligent functions are increasing in the application fields such as smart surveillance, entertainments, sport broadcasting, mobile augmented reality, etc. A large amount of researches to the visual object tracking have treated the issues about robustness to illumination changes, occlusion, background clutter interference and reducing the computational complexity. These researches are generally categorized into 3 methodologies, i.e. point, kernel and silhouette based tracking[1,2]. Among these approaches, the kernel-based tracking has received more interests due to lower computational complexity compared with other methodologies. This efficiency comes from mean-shift procedure included in the kernel-based approach, which makes the tracking position converge to an optical point in fast.

In [3], Comaniciu et al. proposed a kernel-based mean-shift tracker based on Bhattacharyya similarity measure between color histograms of a target and a target candidate. Although kernel-based mean-shift tracking using color similarity measure is generally known to be more robust to partial occlusion than other methodologies, tracking is easy to be failed in the occlusion circumstance. In order to solve the occlusion problem, many approaches[4-9] have been proposed in the last decade. In [4] and [5], authors presented hybrid methods which jointly employed particle filters and mean-shift tracking. The main idea of hybrid methods is to overcome the drawbacks of particle filters which are stable in tracking but require heavy computation, by combining with mean-shift tracking. As different kinds of approaches, methodologies based on partitioning a target region have been proposed[4,6,7,8,9,10,11]. However, these approaches generally include the processing overhead to combine heterogeneous tracking methods. In [4], Khan et al. introduced multi-mode anisotropic mean-shift through partitioning a rectangular bounding box for a target. Jeyakar et al. proposed a weighted fragment based approach that tackled partial occlusion, in which the weights were derived from the difference between the fragment and background colors[6]. In their research, the fragments of a target should be divided manually and a target model for each fragment was built by taking its color histogram after centering an Epanechnikov kernel in it. Zhang et al. [7] proposed a novel local descriptor named local color texture pattern(LCTP), to model the appearance of the object with color and texture information simultaneously. Furthermore, they divided target into multiple blocks and then represented each block with LCTP histogram. Occluded blocks were discarded in computing similarity between target and target candidate.

In this paper, we present a novel kernel-based mean-shift tracking which is robust in partial occlusion by applying multi-mode local kernel weight. The proposed method does not require to manually define fragments of a target as shown in the previous fragment-based approaches and splits a rectangular box for a target into non-overlapping sub-regions of a same aspect ratio as proposed in [4]. In our method, sub-kernel for each sub-region is defined with kernel weight to be determined according to the location of the sub-kernel. In order to validate enhancement of the tracking robustness, we compared tracking errors of our method with those of the previous multi-mode anisotropic mean-shift method without local kernel weight[4].

The rest of the paper is organized as follows. Section 2 explains the concept of mean-shift object tracking. Section 3 describes multi-mode mean-shift tracking with local kernel weight which we propose in this paper. Experiments for the comparison study and their results are presented in Section 4. Finally, we conclude this paper in Section 5.

2. Mean-shift Object Tracking

2.1. Target Model Representation

The kernel-based mean-shift tracking is used to track an ellipsoidal region of a target object based on kernel weighted histogram of colors[3]. In an initial frame, a target model $q=\{q_u\}_{u=1..m}$ with background weight is defined by the probability of the feature u=1..m computed as

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

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

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

v_u is a background weight for background interference reduction. v_u is computed using background histogram $\{o_u\}_{u=1\ldots m}$ (with $\sum_{u=1}^{m} o_u = 1$) and its smallest nonzero entry o^*.This representation is computed in a region around the target. We used a background area equal to three times the target area. The weights v_u is defined by

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

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 the following 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

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

where

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

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).

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

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

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

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

3. Multi-mode mean-shift tracking

3.1. Topology for Partitioning the Target Region

In our method, the topology for partitioning a target region into sub-regions proposed in [4] was applied. The bounding box for a target object is partitioned into non-overlapping sub-regions of a same aspect ratio, i.e., let the bounding box of a candidate object be partitioned into M sub-regions, R^i, i=1,…,M. Any two partitioned sub-regions satisfy $R^i \cap R^j = \emptyset$ if $i \neq j$ and $\cup_i R^i = R$. We assume that all disjoint sub-regions have the same aspect ratio and partitioned sub-regions are non-symmetric. This constraint is to simplify the kernel bandwidth estimate and to automatically divide a target region without the manual operation. Partitioning a target region into smaller sub-regions would allow the mean shift to search several modes.

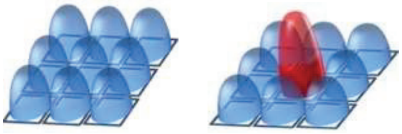
3.2. Target Model with Multi-mode local kernel

In this subsection, we describe the object similarity metric based on the Bhattacharyya coefficient that is applied in the partitioned sub-regions, using spatial kernel-weighted color histograms and multi-mode local kernel weight.

In our method, we define nine non-overlapping sub-regions of a same aspect ratio which are rectangular in shape. A target model $q = \{q^i\}_{i=1..9}$ is defined by the probability of the feature u=1..m for sub-region $q^i = \{q_u^i\}_{u=1..m}$ (with $\sum_{u=1}^{m} q_u^i = 1$). The probability q_u^i for sub-region R^i is analogy to Eq. (1) except using local kernel weight as

$$q_u^i = C^i v_u r^i \sum_{x_j \in R^i} k\left(\left\|\frac{y_c^i - x_j}{h^i}\right\|^2\right) \delta[b(x_j) - u] , \tag{8}$$

where v_u is a background weight computed by Eq. (3) and C^i is the normalizing constant for the color histogram q_u^i. y_c^i is a center pixel location of sub-region R^i . r^i is 1.0 if the center of R^i is the same with the center of the target region, otherwise, 0.5. r^j is designed based on that the primary parts of the target object locate in the center of the kernel. $k(x)$ is a kernel with Epanechnikov profile, $k(x) = \frac{3}{4}(1 - x^2)$, for $|x| < 1$. h^i is bandwidth of a kernel profile for sub-region R^i. In the Khan's method [4], q_u^i did not include v_u and r^j. In Eq. (8), background weight v_u reduces the background interference which was validated in [12]. Figure 1 represents the difference between sub-kernels in Khan's method and ours.



(a) Khan's kernel (b) The proposed kernel

Figure 1. Comparison of multi-mode sub-kernels

3.3. Target Candidate with Multi-mode local kernel

A target candidate $p = \{p^i\}_{i=1..9}$ is defined by the probability of the feature u=1..m for sub-region $p^i = \{p_u^i\}_{u=1..m}$ (with $\sum_{u=1}^{m} p_u^i = 1$). The probability p_u^i for sub-region R^i is computed as

$$p_u^i(y_0^i) = C^i r^i \sum_{x_j \in R^i} k\left(\left\|\frac{y_0^i - x_j}{h^i}\right\|^2\right) \delta\left[b(x_j) - u\right], \qquad (9)$$

where C^i is the normalizing constant for the color histogram p_u^i. y_0^i and h^i are a center pixel location and kernel bandwidth of sub-region R^i, respectively. r^i is the same with the case of Eq. (8). Note that background weight v_u is not included in computing probability p_u^i.

3.4. Multi-mode Mean-shift Tracking

The search for the new target location y_1 in the current frame is repeatedly computed as Eq. (10) from the initial estimated location y_0 .

$$y_1 = \frac{\sum_{i=1}^{9} \sum_{x_j \in R^i} x_j w_j^i g(\left\|\frac{y_0^i - x_j}{h^i}\right\|^2)}{\sum_{i=1}^{9} \sum_{x_j \in R^i} x_j w_j^i g(\left\|\frac{y_0^i - x_j}{h^i}\right\|^2)}, \qquad (10)$$

where $g(x) = -k'(x)$ and

$$w_j^i = \sum_{u=1}^{m} \delta\left[b(x_j) - u\right] \sqrt{\frac{q_u^i}{p_u^i(y_0^i)}}. \qquad (11)$$

If $\|y_1 - y_0\| < \varepsilon$ or the number of mean-shift iterations comes to N_{max}, typically taken equal to 20, y_1 is determined as the center of the most similar region to the target object. Otherwise, y_1 is assigned to y_0, and nine non-overlapping sub-regions and centers of sub-regions are redefined. Then y_1 is recomputed using Eq. (10).

4. Experimental Results

In order to show the improvement of stability of the proposed method, we implemented the previous multi-mode mean-shift tracking which was proposed in [4] for occlusion handling. We investigated the tracking errors of [4]'s method and ours. The tracking error is defined by the Euclidean distance from the center of ground truth object to the tracked kernel center for each video frame.

To conduct the comparative tests, we selected four video sequences which have characteristic features in camera moving condition, background clutter and the occlusion of a target object. Three of the selected video (Egtest04, Visor1 and Woman) are publicly available and can be found from the PETS[1](Performance Evaluation of Tracking System), Video surveillance Online Repository[2] and the site of Technion-Israel Institute of Technology[3]. Test sequence Tiger 1 was produced by authors using Microsoft LifeCam Web Camera. The properties of each test video sequence are summarized in Table 1 and each representative image is illustrated in Figure 2. In Egtest05 sequence, camera is moving while tracking a target car and it is sometimes occluded by trees. Furthermore, illumination changes among frames are happened. Tiger 1 sequence includes camera movement and full occlusion of a target object. The background is more or less simple. In Visor 1 sequence, camera is static and partial occlusion of a target object is included. Background is not severely complex. Finally, in the case of woman test sequence, a camera is moving and a target woman is sometimes occluded by the parking cars. The background is rather complex and illumination changes are shown.

Table 1. Features of selected video sequences

Name	Resolution	Condition			Availability
		Camera	Occlusion	Background	
Egtest05	640 x 480	moving	full/ partial	clutter	PETS2005
Tiger 1	640 x 480	moving	full	normal	
Visor1	352 x 288	static	partial	normal	openvisor

[1] http://www.cvg.cs.rdg.ac.uk/PETS2005/

[2] http://www.openvisor.org/

[3] http://www.cs.technion.ac.il/~amita/fragtrack/fragtrack.htm/

Woman	352 x 288	moving	partial	clutter	Technion Institute

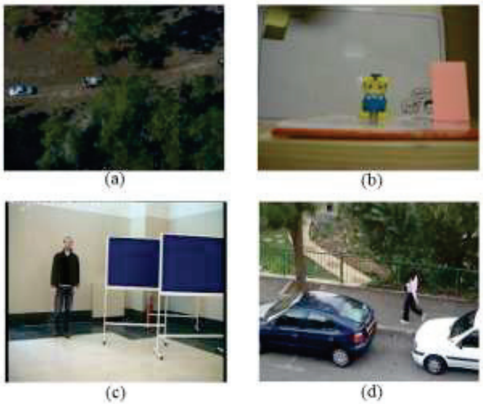


Figure 2. Representative image for the test sequence of (a) Egtest94, (b)Tiger1, (c)Visor1 and (d) Woman.

Table 2 compares tracking errors of the previous multi-mode method[4] and ours. Moreover, the accuracy improvement ratios after applying our method are shown. The accuracy improvement(AI) ratio was computed by

$$AI = \frac{E_{prev} - E_{ours}}{E_{ours}} \times 100 \,, \tag{12}$$

where E_{prev} and E_{ours} are the average tracking errors of the previous multi-mode method[4] and ours, respectively. In the cases of Egtest05 and Woman test sequences, the previous method that did not apply background weight v_u and local sub-kernel weight r^i failed to track a target after background change and partial occlusion. In the comparative test using the selected four test sequences, the accuracy of the proposed method was improved by the average 231.48%

Figure 3 and Figure 4 show comparative test results using Egtest05 and Woman test sequences. In the case of Egtest05, target objects are frequently occluded by trees and background of target object is severely changed. Figure 3 shows that the previous method fails to track a car right after partial occlusion by trees and the proposed method tracks a car continuously even though partial occlusion to the target car is happened. In Figure 4, the previous method misses the target when background change and partial occlusion happen together. Average execution time for tracking per frame using four test sequences was 2.1 msec and 2.3 msec in Intel Core i7 2.67GHz CPU and 4GB RAM by the previous method and our method, respectively. Therefore, the trackers tracked a target within about 50 frames per second.

Table 2. Tracking errors and accuracy improvement

	Egtest05	Tiger1	Visor1	Woman
Previous Method[4]	71.28 [fail]	34.89	11.90	139.0 [fail]
Ours	9.70	21.17	9.81	45.58
Accuracy Improvement	634.85	64.81	21.30	204.96

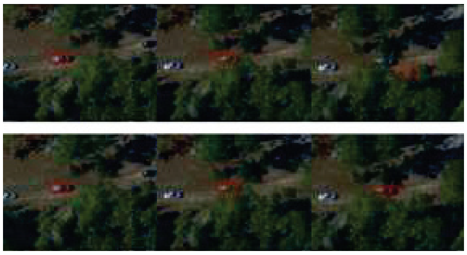


Figure 3. Experimental results of the previous method (upper row) and the proposed method (lower row) using Egtest05 test sequence. From left to right, the frames 82, 122 and 147.

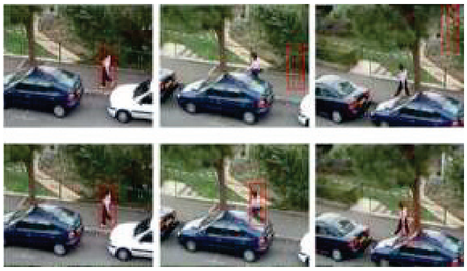


Figure 4. Experimental results of the previous method (upper row) and the proposed method (lower row) using Woman test sequence. From left to right, the frames 174, 204 and 239.

5. Conclusions

Kernel-based mean-shift tracking is known to be more efficient and more stable to partial occlusion than other methodologies. However, tracking is still failed easily in the cases of partial or full occlusion to a target object. For more efficient and robust performance of kernel-based mean-shift tracking in partial occlusion cases, this paper proposed a mean-shift tracking with multi-mode local kernel weight.

This paper was based on the multi-mode mean-shift approach that was proposed in [4], and improved robustness compared to it. In order to define sub-kernels, we first splits a rectangular box for a target into non-overlapping sub-regions of a same aspect ratio. Therefore our method does not require to manually define fragments of a target as shown in the previous fragment-based approaches. In our method, sub-kernel for each sub-region is defined with kernel weight to be determined according to the location of the sub-kernel. Furthermore, the background weight for background clutter reduction is considered in target model construction.

Using the comparative experiments, we validated that the proposed method is more stable than the previous multi-mode mean-shift approach in partial occlusion and background clutter. In experiments, the previous approach failed to track a target in two sequences among four selected test sequences while our method completed to track targets in all test sequences. We measured tracking errors of comparative methods using ground truth data and computed accuracy improvement ratio. These experiments showed that the accuracy of the proposed method was improved by the average 231.48%.

Full occlusion issue is beyond this paper and this problem can not be solved by using only kernel-based mean-shift approach. Therefore, as a future work, we would like to study about occlusion level determination based on multi-mode mean-shift approach and full occlusion handling approach after decision of high-level occlusion.

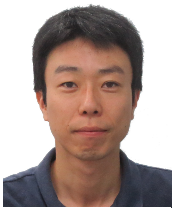
Acknowledgment

This research was partly supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2014-H0301-14-1012) supervised by the NIPA (National IT Industry Promotion Agency), and partly supported by the International Collaborative R&D Program of the Ministry of Knowledge Economy(MKE), as a result of project (2010-TD-300802-002)

References

1. A. Yilmaz, O. Javed, M. Shah. Object tracking: a survey, ACM Computing Surveys, 2006:38(4): 45, Article 13.
2. M. J. Patel, B. Bhatt. A comparative study of object tracking technoques, International Journal of Innovative Research in Science, Engineering and Technolgy, 2015: 4(3): 1361-1364.
3. D. Comaniciu, V. Ramesh, P. Meer. Kernel-based object tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence 2003:25(5):564-577.
4. Z. H. Khan, I. Y.-H. Gu, A. G. Backhouse. Robust Visual Object Tracking using Multi-Mode Anisotropic Mean Shift and Particle Filters, IEEE Transactions on Circuits and Systems for Video Technology 2011:21(1):74-87.
5. S. Zhang, Robust visual tracking based on occlusion detection and particle redistribution, ICIMCS 2010:159-162.
6. J. Jeyakar, R. V. Babu, K.R. Ramakrishnan, “Robust object tracking with background-weighted local kernels”, Computer-Vision and Image Understanding 2008:112:296-309.
7. S. Zhang, H. Yao, S. Liu. Partial occlusion robust object tracking using an effective appearance model, Visual Communications and Image Processing 2010:1-8.
8. A. Adam, E. Rivlin, I. Shimshoni. Robust fragments-based tracking using the integral histogram, Proc. CVPR, 2006:1:798-805.
9. Z. Fan, Y. Wu, M. Yang. Multiple collaborative kernel tracking, Proc. CVPR, 2005:2:502-509.
10. V. Rowghanian, K. Ansari-Asl. Object tracking by mean shift and radial basis function neural networks, 2015:1-18.
11. D. Jia, L. Zhang, C. Li. The improvement of mean-shift algorithm in target tracking, International Journal of Security and Its Applications, 2015:9(2):21-28.
12. J. Ning, L. Zhang, D. Zhang, C. Wu. Robust mean-shift tracking with corrected background-weighted histogram. IET Comput. Vis, 2012:6:62-69.

〈저자소개〉



김은섭

- 2000년 경희대학교 기계공학과 학사
- 2012년 KGIT 뉴미디어학부 석사
- 2012년 ~ 현재 KETI 멀티미디어IP연구센터 위촉연구원
- 관심분야 : 비주얼 트래킹, 머신비젼, 병렬 컴퓨팅

김용구

- 1993년 연세대학교 전기공학과 학사
- 1995년 연세대학교 전기공학과 석사
- 2001년 연세대학교 전기전자공학과 박사
- 2009년~현재 한독미디어대학원대학교 뉴미디어콘텐츠학과 교수
- 관심분야: 영상신호처리, 비디오압축, 디지털방송시스템

최유주

- 1989년 이화여자대학교 전자계산학과 학사
- 1991년 이화여자대학교 전자계산학과 석사
- 2005년 이화여자대학교 컴퓨터공학과 박사
- 2010년 ~ 현재 한독미디어대학원대학교 뉴미디어콘텐츠학과 부교수
- 관심분야 : 컴퓨터 그래픽스, 영상처리, HCI, 모바일 증강현실