

# 다중 스케일 커널화 상관 필터를 이용한 견실한 객체 추적

## Robust Object Tracking based on Kernelized Correlation Filter with multiple scale scheme

윤 준 한\*, 김 진 현\*

Jun Han Yoon\*, Jin Heon Kim\*

### Abstract

The kernelized correlation filter algorithm yielded meaningful results in accuracy for object tracking. However, because of the use of a fixed size template, we could not cope with the scale change of the tracking object. In this paper, we propose a method to track objects by finding the best scale for each frame using correlation filtering response values in multi-scale using nearest neighbor interpolation and Gaussian normalization. The scale values of the next frame are updated using the optimal scale value of the previous frame and the optimal scale value of the next frame is found again. For the accuracy comparison, the validity of the proposed method is verified by using the VOT2014 data used in the existing kernelized correlation filter algorithm.

### 요 약

커널 상관 필터 알고리즘은 객체 추적에 대해 정확도에서 의미 있는 성과를 거두었다. 그러나 고정된 크기의 템플릿을 사용하기 때문에 추적 대상의 스케일 변화에 대처할 수 없었다. 본 논문에서는 최근접 보간법과 표준 가우시안 정규화를 이용한 다중 스케일에서의 상관 필터링 응답 값을 이용하여 프레임별로 가장 적합한 스케일을 찾아 객체를 추적하는 방식을 제안한다. 다음 프레임의 스케일 값들은 이전 프레임의 최적 스케일 값을 이용해 갱신하고 다시 해당 프레임에서의 최적의 스케일 값을 찾는다. 정확도 비교를 위해 기존 커널 상관 필터 알고리즘에서 사용된 VOT2014 데이터를 사용하여 제안된 방법의 유효성을 검증한다.

*Key words* : Visual tracking, Correlation filters, Gaussian distribution, Computer vision, HOG

### 1. 서론

휴먼 컴퓨터 인터페이스, 동작 분석, 감시 등의 응용 애플리케이션에서 고전적이면서 가장 일반적으로 발생하는 문제 중의 하나가 시각적 객체 추적의 정확성이다. 시각적 객체 추적이란 비디오의 각 프레임에서 초기화된 시각적 대상의 위치를 추정하는 것이다. 최근에 제안된 여러 객체 추적 방법

[1][2][3][4] 들도 조명 변화, 모션 블러, 회전 및 카메라 모션 등을 포함한 몇 가지 문제로 인해 제 성능을 발휘하지 못하고 있다.

최근 상관 필터(Correlation Filter)가 적용된 여러 추적 응용 애플리케이션들이 등장하고 있다[5][6][7][8]. 상관 필터는 신호처리 분야에서 두 신호 사이의 유사성 척도로 쓰일 수 있는 메트릭을 제공하고, 연산량이 많은 공간계열에서의 상관 연산 대

\* Dept. of Computer Engineering, Seokyeong University

★ Corresponding author

E-mail : jinheon@skuniv.ac.kr, Tel : +82-2-940-7747

※ Acknowledgment

This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program 2017.

Manuscript received Sep. 10, 2018; revised Sep. 19, 2018; accepted Sep. 20, 2018

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

신 푸리에 영역에서의 적은 연산으로 대체할 수 있는 이점이 있다.

이를 통해 *Bolme et al.*[9]과 *Henriques et al.*[10]은 상관 필터를 객체 추적에 응용하여 도입하였고, 그 산물인 KCF(Kernelized Correlation Filter) 알고리즘[23][24][27]은 VOT2014 challenge [11]에서 우수한 성능을 입증함으로써 객체 추적 분야에 있어 상당한 진보를 꾀하였다. KCF는 객체의 모양 측면에서 학습에 사용한 객체 데이터를 닮은 대상과의 상관 필터링 응답을 최대화하는 최상의 필터 탭을 찾는다. 이때 오버샘플링 전략으로 찾고자 하는 대상 이미지에 대한 무작위 표본들을 학습하는 대신 대상 이미지에서의 윈도우 내 가능한 모든 변환을 고려한다. 이 방법은 많은 수의 중복 표본들이 있어야 하므로 이전에는 단점으로 간주하였으나 이들이 적절하게 구성될 시 대상 이미지를 잘 표현하는 순환 행렬을 형성하게 되며, 이에 대한 대각화는 DFT(Discrete Fourier Transform) 행렬을 사용할 시 효율적으로 계산될 수 있다는 이점을 갖게 된다. 따라서 오버샘플링 전략을 사용하여 DFT 대각화 후, 고정회귀문제(Ridge Regression)를 주파수 도메인에서 해결함으로써 적은 연산량으로 높은 정확도의 객체 추적이 가능해졌다.

이와 같은 이점들로 인해 찾고자 하는 대상의 위치를 실시간으로 비교적 정확하게 예측할 수 있게 되었으나, 고정 크기의 템플릿을 사용함으로써 인해 찾고자 하는 대상 이미지의 스케일 변화에 대처할 수 없었다.

본 논문에서는 여러 스케일링 값들에 대한 탐지 결과를 표준 가우시안 분포(Standard Gaussian Distribution)를 이용하여 정규화한 후, 가장 우수한 값에 해당하는 스케일링 값으로 갱신 후 적용하여 이를 해결한다.

## II. 관련 연구

일반적으로 추적 알고리즘은 두 가지 주요 범주로 나눌 수 있다. 첫 번째로 대상 관측을 설명하기 위한 모양 모델을 채택하고 이와 모양이 가장 유사한 대상을 찾는 방법론이 있다. 이 유형의 추적 알고리즘의 주요 목표는 대상 이미지가 다양한 변화를 하더라도 이를 안정적으로 설명할 수 있는 모델을 생성하는 것이다. 해당 범주의 알고리즘에는

Mean shift tracker[12], Incremental tracker(IVT)[13], Fragment-based tracker(Frag)[14], L1-min tracker[15], Multi-task tracker(MTT)[16], Lowrank sparse tracker[17], 그리고 Structural sparse tracking [18] 등이 있다.

두 번째로 시각적 객체 추적을 배경과 가장 다른 대상 위치를 검색하는 분류 문제로 해결하는 방법론이 있다. 해당 범주의 예로는 Multiple instance learning tracking(MIL)[19], Ensemble tracking[20], Support vector tracking[21] 및 상관관계를 고려한 추적 알고리즘 등이 있다. 상관 필터는 분류 문제에서 오랫동안 사용됐는데, 필터의 에너지 응답을 최소화하는 필터 탭을 학습하거나 주어진 학습 데이터에 대한 응답의 분산 값을 최소화하는 것[22]을 목적으로 한다.

이후 객체 추적을 위해 상관 필터를 사용하는 사례가 등장하기 시작했다. 예를 들어, 여러 개의 KCF 추적 알고리즘을 사용하여 객체의 다른 부분을 나타내고 이를 함께 추적하거나[25], 논문 [26]에서와 같이 더 큰 영역에서 온라인 무작위 Fern 분류자를 학습하여 장기간의 가림 현상이나 표적이 시야 밖으로 나가면 실패를 식별하고 객체를 다시 탐지하는 등의 알고리즘 등이 있다.

## III. 제안하는 방법

KCF의 가장 큰 결점은 고정 크기의 템플릿 사용으로 인해 스케일 변화에 대한 대처가 미흡하다는 것이다. 본 논문에서는 프레임마다 여러 스케일별로 대상 이미지 패치에 최근접 보간법(Nearest neighbor interpolation)을 적용하고 상관 필터링 응답 값을 구한다. 이후 각 스케일에 따른 응답 값들에 대해 가우시안 분포를 적용 후 최대 응답 값에 해당하는 적정 스케일을 찾아 기존 KCF 알고리즘의 성능을 향상하고자 한다. 제안한 알고리즘의 적용 과정은 다음과 같다.

### 1. 초기 스케일 후보 값 정의

스케일 값 1을 중심으로 프레임  $t$ 에서의 임의의 스케일 값  $t_n$ 들을 식 (1)과 같이 정의한다.

$$S_t = (t_1, t_2, \dots, 1, \dots, t_{n-1}, t_n) \quad (1)$$

## 2. 스케일별 이미지 크기 계산

대상 이미지 패치의 크기를 나타내는 *patch size*에 스케일  $S_t$ 를 적용하여  $t_n$ 에 따른 대상 이미지 패치의 크기  $u_n$ 을 다음과 같이 계산한다.

$$u_n = t_n \times \text{patch size} \quad (2)$$

## 3. 최근접 보간법 적용

최근접 보간법은 가장 가까운 위치에 있는 픽셀의 값을 참조하려는 방법이다. 예를 들어 보간법이 역방향으로 맵핑 시 원본 영상의 참조 좌표가 (28.2, 40.8)로 계산되었다면, 이 위치에서 가장 가까운 정수 좌표인(28, 41)의 픽셀값을 그대로 사용하는 방식이다. 구현이 간단하고 연산량이 적기 때문에 실시간 구현 측면에서 처리속도에 민감한 추적 알고리즘에 적용하기 쉽다.

앞서 계산한 스케일별 대상 이미지 패치 크기  $u_n$ 에 따라 원본 이미지 패치에 최근접 보간법을 적용한다. 이를 통해 스케일 적용을 통해 변환된 이미지 패치에서 상관 필터링 응답 값 계산에 사용될 특징을 좀 더 정확하게 추출할 수 있다.

## 4. 상관 필터링 응답 값 계산

최근접 보간법이 적용된 다양한 스케일별 이미지 패치들에 대해 HOG(Histogram of Oriented Gradients) 특징을 추출한 후 푸리에 변환을 한다. 이후 해당 값들과 식 (3) 간의 상관 필터링 응답 값  $r_n$ 을 식 (4)를 이용해 구한다[27].

$$\hat{\alpha}^* = \frac{\hat{y}}{\hat{k}^{xx} + \lambda} \quad (3)$$

$$\hat{f}(z) = (\hat{k}^{xz})^* \odot \hat{\alpha} \quad (4)$$

$$R = (r_1, r_2, \dots, r_n) \quad (5)$$

## 5. 표준 가우시안 정규화

식 (5)의 상관 필터링 응답 값  $R$ 을 식 (6)을 통해 정규화한다.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (6)$$

표준 가우시안 정규화로 인해 스케일 갱신 과정에서 이전 프레임에 적용된 스케일링 값에 이어 더욱 자연스러운 변화로 다음 프레임의 스케일링 값을 정할 수 있다.

## 6. 최대 응답 값 도출 및 스케일 후보 갱신

정규화된 응답 값 중 가장 큰 값을 구하고, 이에 해당하는 스케일을 BS(Best Scale)로 설정한다.  $S_{t_{mid}}$ 는 프레임 t에서의 스케일 후보 값에서 중간에 해당하는 값이다. 다음 프레임의 스케일 후보  $S_{t+1}$ 은 아래와 같이 갱신된다.

$$S_{t+1} = S_t \times \frac{BS}{S_{t_{mid}}} \quad (7)$$

지금까지의 알고리즘 과정은 다음의 그림 1의 알고리즘 처리도를 통해 요약할 수 있다.

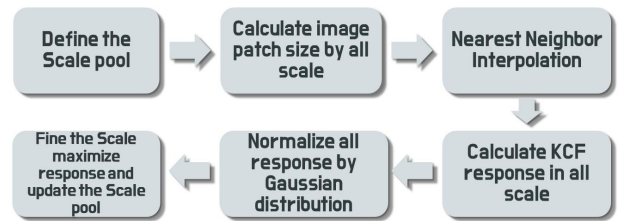


Fig. 1. Flowchart for Multi-Scale KCF.

그림 1. 다중 스케일 KCF 처리도

## IV. 실험 방법 및 결과 분석

추적 알고리즘의 성능 비교 시 가장 많이 사용되는 VOT2014 challenge 데이터를 이용하였다. 이미지 시퀀스로 제공되는 총 25개의 영상이 있으며, 부분 가림, 조명 변화, 모션 블러 등이 반영되어 있다. 같은 실험 환경에서 기존의 추적 알고리즘인 Struck, MIL, IVT, KCF의 성능과 비교하였다.

### 1. 모델 파라미터 설정

이미지 패치 특징은 가우시안 커널 함수를 사용한 HoG(Histogram of Gaussian)를 사용하였으며, 기존의 KCF 알고리즘 실험 환경과 같도록 가우시안 커널의  $\sigma$ 는 0.5, HoG의 셀 크기는 4x4, orientation bins의 값은 9로 설정하였다. 추출된 특징 벡터는 Hanning window를 이용해 가공하였으며, 초기 스케일 값은 [0.95, 0.96, 0.98, 1, 1.02, 1.03, 1.05]로 설정하였다.

### 2. 컴퓨팅 환경

본 논문에서 제안한 실험은 Intel(R) Core(TM) i7-7700 3.60GHz CPU와 16GB 메모리가 탑재된 PC에서 수행하였다.

3. 성능 평가 방법

성능 평가 실험에는 두 가지 평가 기준을 사용하였다. 첫 번째는 CLE(Center Location Error)로, 중심 위치 차이 값들의 평균이다. 추적된 결과의 중심 위치와 실제 중심 위치의 차이 값이 작을수록 더 정확함을 의미한다. 두 번째는 파스칼 VOC 중첩 비인 VOR(Voc Overlap Ratio)[28] 로 다음과 같이 정의된다.

$$VOR = \frac{area(ROI_T \cap ROI_G)}{area(ROI_T \cup ROI_G)} \quad (8)$$

여기서  $ROI_T$ 는 tracking bounding box를,  $ROI_G$ 는 ground truth bounding box를 의미하며 VOR 수치가 높을수록 높은 정확도를 나타낸다.

알고리즘별로 전체 25개 영상에 대한 평균 VOR 값과 평균 CLE 값을 비교하여 전반적인 성능을 비교하였고, 객체의 부분 가림(Occlusion) 영상들과 조명 변화, 모션 블러 등(Deformation)의 현상이 나타나는 영상들에 대해서도 같은 방법으로 성능을 비교하였다.

4. 결과 및 분석

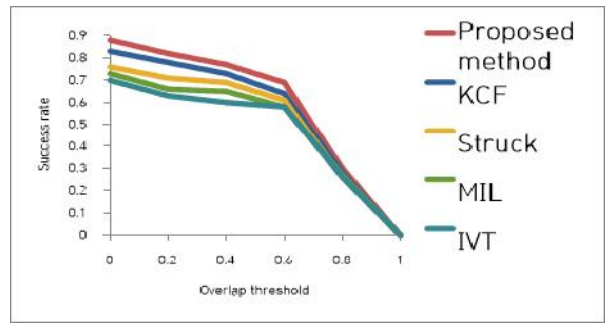
표 1은 제안된 추적 알고리즘과 다른 4개 추적 알고리즘의 전반적인 성능을 보여 준다. 제안된 접근 방식은 평균 VOR에서 56.2%를 달성하였다. 이는 기존 KCF 보다 약 10% 향상되었음을 의미한다.

Table 1. Comparison with state-of-the-art trackers.

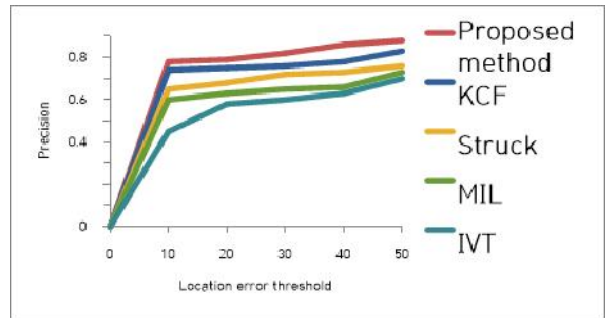
표 1. 우수한 추적 알고리즘들과의 전반적 성능 비교

	Proposed method	KCF	Struck	MIL	IVT
mean VOR	0.562	0.513	0.456	0.423	0.395
mean CLE	29.98	34.47	49.92	52.04	54.63

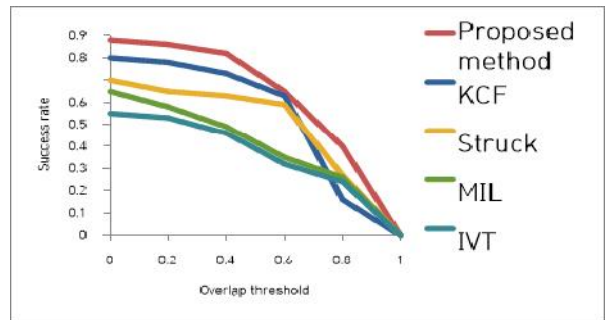
그림 2에서 (a), (b)는 전체 영상에 대해 알고리즘별로 각각 VOR, CLE를 통한 정확성 추이를 보여 주고, (c), (d)는 조명 변화, 모션 블러 등의 영상들, (e), (f)는 부분 가림 영상들에 대해 각각 VOR, CLE를 이용한 정확성 추이를 나타낸다. 객체 추적 시 거론되는 대표적인 상황들에 대해 제안된 알고리즘의 성능이 앞서는 것을 확인할 수 있다.



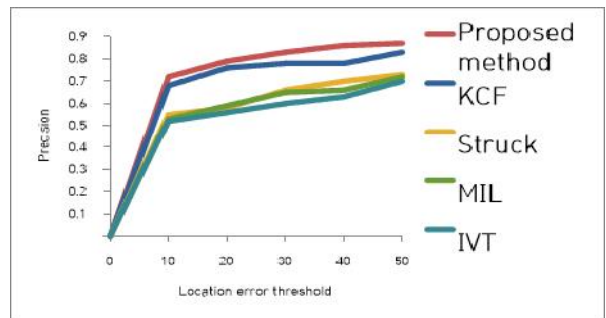
(a)



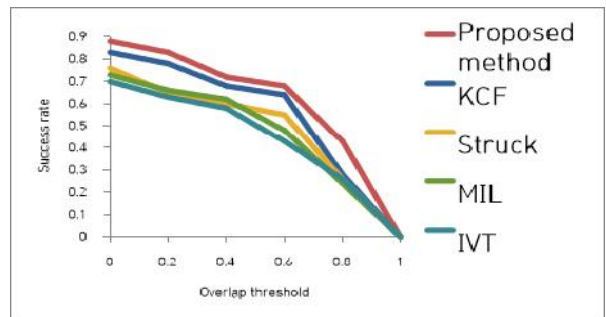
(b)



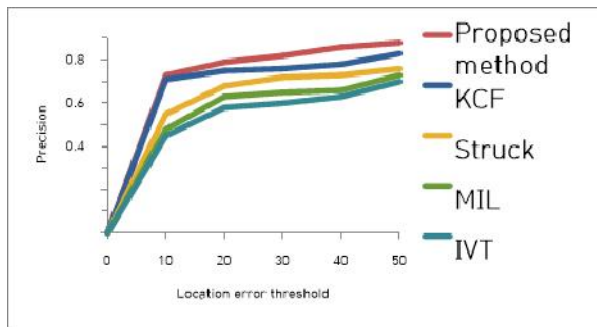
(c)



(d)



(e)



(f)

Fig. 2. The graph for the proposed method compared with 4 state-of-the-art trackers. (a)~(f) indicate the VOR and CLE of overall, deformation, occlusion.

그림 2. 4개의 우수한 추적 알고리즘과 비교한 그래프. (a)~(f)는 차례대로 전체 영상, 조명 변화 및 모션블러 영상, 부분 가림 영상들의 VOR, CLE 정확성의 변화를 보여 준다.

## V. 결론

본 논문에서는 상관 필터의 프레임워크를 기반으로 한 효과적인 추적 알고리즘을 제시하였다. 기존의 KCF에서는 고정된 크기의 템플릿 사용으로 객체의 스케일 변화에 대한 대처가 미흡하였다. 제안된 알고리즘은 최근접 보간법과 가우시안 정규화를 이용한 다중 스케일 방식을 적용하여 이를 개선하였고, 다양한 객체 추적 시나리오가 담겨 있는 VOT 2014 데이터를 통해 이를 입증하였다.

## References

- [1] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *Acm computing surveys (CSUR)*, vol.38, no.4, pp.13, 2006. DOI:10.1145/1177352.1177355
- [2] S. Salti, A. Cavallaro, and L. D. Stefano, "Adaptive appearance modeling for video tracking: Survey and evaluation," *Image Processing, IEEE Transactions on*, vol.21, no.10, pp.4334-4348, 2012. DOI:10.1109/TIP.2012.2206035
- [3] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark. In Computer vision and pattern recognition," *CVPR, 2013 IEEE Conference on*, pp.2411-2418. 2013.
- [4] A. Smeulders, D. Chu, R. Cucchiara, S.

Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.36, no.7, pp.1442-1468, 2014. DOI:10.1109/TPAMI.2013.230

[5] Boddeti, V.N., Kanade, T., Kumar, B.V.: "Correlation filters for object alignment," *CVPR 2013 IEEE Conference on*, pp.2291-2298, 2013. DOI:10.1109/CVPR.2013.297

[6] Galoogahi, H.K., Sim, T., Lucey, S. "Multi-channel correlation filters," *ICCV*, pp.4321-4328, 2013. DOI:10.1109/ICCV.2013.381

[7] Henriques, J. F., Carreira, J., Caseiro, R., Batista, J, "Beyond hard negative mining: Efficient detector learning via block-circulant decomposition," *ICCV*, pp.2760-2767, 2013. DOI:10.1109/ICCV.2013.343

[8] Revaud, J., Douze, M., Cordelia, S., Jgou, H, "Event retrieval in large video collections with circulant temporal encoding," *CVPR*, 2013, pp. 2459-2466.

[9] D. S. Bolme, J. R. Beveridge, B. A. Draper, Lui, Y. M, "Visual object tracking using adaptive correlation filters," *CVPR*, pp.2544-2550, 2010. DOI:10.1109/CVPR.2010.5539960

[10] Henriques, J. F., Carreira, J., Caseiro, R., Batista, J, "Beyond hard negative mining: Efficient detector learning via block-circulant decomposition," *ICCV*, pp.2760-2767, 2013. DOI:10.1109/ICCV.2013.343

[11] M. Kristan, R. Pflugfelder, A. Leonardis, J. Matas, L. Cehovin, G. Nebehay, T. Vojř í r, G. Fernandez, A. Lukezić, A. Dimitriev, et al, "The visual object tracking vot 2014 challenge results," *Computer Vision-ECCV 2014 Workshops*, 2014, pp.191-217.

[12] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking. Pattern Analysis and Machine Intelligence," *IEEE Transactions on*, vol.25, no.5, pp.564 - 577, 2003. DOI:10.1109/TPAMI.2003.1195991

[13] T. Poggio and G. Cauwenberghs, "Incremental and decremental support vector machine learning," *Advances in neural information processing systems*,



pp.409, 2001.

- [14] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragmentsbased tracking using the integral histogram," *Computer vision and pattern recognition, 2006 IEEE Computer Society Conference on*, vol.1, pp.798-805, 2006. DOI:10.1109/CVPR.2006.256
- [15] X. Mei and H. Ling, "Robust visual tracking using L1 minimization," *Computer Vision, 2009 IEEE 12th International Conference on*, pp.1436-1443, 2009. DOI:10.1109/ICCV.2009.5459292
- [16] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via multi-task sparse learning," *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp.2042-2049, 2012. DOI:10.1109/CVPR.2012.6247908
- [17] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Low-rank sparse learning for robust visual tracking," *Computer Vision - ECCV 2012*, pp. 470-484, 2012. DOI:10.1007/978-3-642-33783-3\_34
- [18] T. Zhang, S. L. C. Xu, S. Yan, B. Ghanem, N. Ahuja, and M.-H. Yang, "Structural sparse tracking," *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pp.150 - 158, 2015. DOI:10.1109/CVPR.2015.7298610
- [19] B. Babenko, M.-H. Yang, and S. Belongie, "Visual tracking with online multiple instance learning," *Computer Vision and Pattern Recognition, CVPR 2009. IEEE Conference on*, pp.983-990, 2009. DOI:10.1109/CVPR.2009.5206737
- [20] S. Avidan, "Ensemble tracking," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.29, no.2, pp.261-271, 2007. DOI:10.1109/TPAMI.2007.35
- [21] S. Avidan, "Support vector tracking," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.26, no.8, pp.1064-1072, 2004. DOI:10.1109/TPAMI.2004.53
- [22] B. Vijaya Kumar, "Minimum-variance synthetic discriminant functions," *JOSA A*, vol.3, no.10, pp.1579 - 1584, DOI:1986.10.1364/JOSAA.3.001579
- [23] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," *Computer*

*Vision - ECCV 2012*, pp. 702 - 715, 2012.

DOI:10.1007/978-3-642-33765-9\_50

- [24] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High speed tracking with kernelized correlation filters," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.37, no.3, pp.583 - 596, 2015. DOI:10.1109/TPAMI.2014.2345390
- [25] T. Liu, G. Wang, and Q. Yang, "Real-time part-based visual tracking via adaptive correlation filters," *Intelligence*, pp.2345-2390, 2015. DOI:10.1109/CVPR.2015.7299124
- [26] C. Ma, X. Yang, C. Zhang, and M.-H. Yang, "Long-term correlation tracking," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.5388 - 5396, 2015. DOI:10.1109/CVPR.2015.7299177
- [27] Henriques, J. F., Caseiro, R., Martins, P., Batista, J. "High-speed tracking with kernelized correlation filters," *TPAMI*, vol.37, no.3, pp.583-596, 2015. DOI:10.1109/TPAMI.2014.2345390
- [28] Galoogahi, H. K., Sim, T., Lucey, S, "Multi-channel correlation filters," *ICCV*, pp. 4321-4328, 2013. DOI:10.1109/ICCV.2013.381

## BIOGRAPHY

### Jun Han Yoon (Member)



2017 : BS degree in Computer Engineering, Seokyeong University.  
2017~Present : MS course in Electronics & Computer Engineering, Seokyeong University.

### Jin Heon Kim (Member)



1982 : BS degree in Electrical Engineering, Korea University.  
1984 : MS degree in Electrical Engineering, Korea University.  
1994 : PhD degree in Electrical Engineering, Korea University.

1995~Present : Associate Professor, Seokyeong University.