Unsupervised Moving Object Detection through Background Models for PTZ Camera

Kimin Yun, Hyungil Kim, Kangmin Bae, Jougyoul Park

Introduction

❖ Moving object detection in a PTZ Camera

- Video segmentation, motion clustering, semantic segmentation...
- Background-centric approach
  - Motion compensation + Background subtraction
  - VERY fast, Unsupervised, Suitable for pre-processing

- Problem
  - Naive approach: Many false positive by compensation error

  - Existing problem of conventional approach
    - Foreground Loss problem
    - Background Contamination Problem

Proposed Method

❖ Fine background model

- Extend the Vibe for moving camera
- Spatio-temporal update is applied
- Model initialization and update rule is changed by camera movement

Input — Coarse background — Fine background

$$C_i = \sum_{j \in S_i} \sum_{n=1}^{N} \mathbb{1}(D(I_i, \tilde{v_j}^{(n)}) < R),$$

spatio-temporal update rule

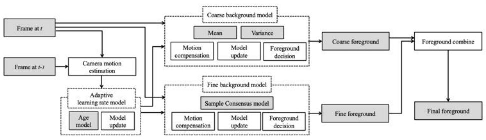
Algorithm 1: Updating the fine background model

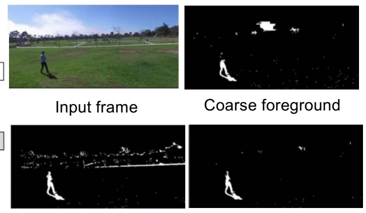
for *Each pixel i on current frame $I^{(t)}$* **do**
- **if** $t = 0$ **then**
  - $\mathbf{M}_i$ is initialized to $\{I_i^{(0)}, I_i^{(0)}, \ldots, I_i^{(0)}\}$
- **else**
  - Motion compensation is applied to $\mathbf{M}_i$.
  - **if** $t < N$ **then**
    - $\tilde{v_i}^{(t)}$ is removed from $\mathbf{M}_i$
    - $I_i^{(t)}$ is inserted to $\mathbf{M}_i$
  - Compute C_i in equation (8)
  - **if** $C_i \geqslant \#_{min}$ **then**
    - $P \leftarrow \min(\alpha_i, \phi)$
    - $p \sim Uniform(0, P-1)$
    - **if** $p = 0$ **then** ▷ update for pixel i
      - $n \sim Uniform(0, N-1)$
      - $\tilde{v_i}^{(n)}$ is removed from $\mathbf{M}_i$
      - $I_i^{(t)}$ is inserted to $\mathbf{M}_i$
    - $p_2 \sim Uniform(0, P-1)$
    - **if** $p_2 = 0$ **then** ▷ update for neighbor pixel j
      - $k \sim Uniform(0, K)$
      - $j \leftarrow S_i(k)$
      - $n \sim Uniform(0, N-1)$
      - $\tilde{v_j}^{(n)}$ is removed from $\mathbf{M}_j$
      - $I_i^{(t)}$ is inserted to $\mathbf{M}_j$

camera movement handling

model initialize & update

❖ Combine Two backgrounds characteristics



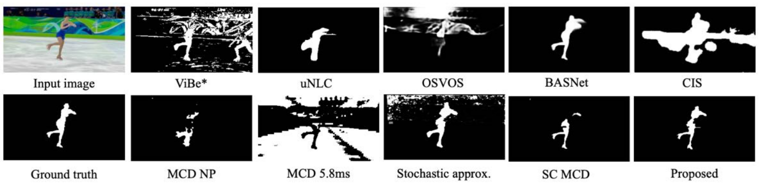


Input frame — Coarse foreground — Fine foreground — Final foreground

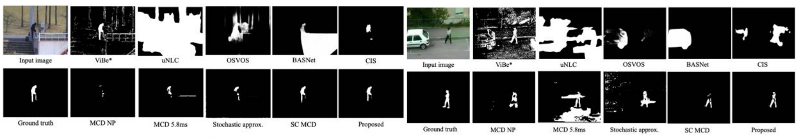
Experiments

❖ F-measure of Moving camera dataset

Method	walking	skating	woman	woman2	fence	ground1	ground2	ground3	ground4	ground5	average
ViBe* [6]	0.0375	0.2229	0.0375	0.0929	0.1042	0.5656	0.4733	0.4118	0.0299	0.1309	0.2107
FIC* [8]	0.0613	0.2373	0.0361	0.1345	0.0954	0.4543	0.4108	0.1538	0.0453	0.1319	0.1761
BMRI-ViBe* [9]	0.0438	0.2402	0.0400	0.0921	0.1104	0.4249	0.3868	0.2161	0.0383	0.1377	0.1730
MCD NP [25]	0.4351	0.4164	0.4935	0.5791	0.2691	0.2773	0.3750	0.1222	0.1969	0.3540	0.3519
MCD 5.8ms [26]	0.7349	0.2447	0.3395	0.3448	0.7357	0.6573	0.7177	0.1531	0.5274	0.0678	0.4523
Stochastic approx [28]	**0.8335**	0.6543	0.3986	**0.8783**	**0.8788**	0.2221	0.2792	0.0181	0.0111	0.2181	0.4392
FP Sampling [27]	0.7058	0.8539	0.7268	0.5828	0.7654	0.7977	0.8306	0.1396	0.4226	0.8212	0.6646
SC MCD [29]	0.7496	0.8560	0.6650	0.6311	0.7637	0.8965	**0.9118**	0.8843	0.8824	0.9326	0.8173
uNLC [32]	0.0158	0.1419	0.0178	0.0487	0.0346	0.0570	0.0342	0.0216	0.0031	0.0143	0.0389
OSVOS [1]	0.3397	0.5344	0.0121	0.1260	0.7033	0.7697	0.5447	**0.9696**	0.0050	0.1224	0.4127
CIS [33]	0.0538	0.3036	0.1522	0.4681	0.1180	0.1545	0.0862	0.0581	0.0046	0.0184	0.1418
BASNet [34]	0.3433	0.9379	0.0205	0.2289	0.2119	0.6039	0.9564	0.9586	**0.9439**	**0.9829**	0.6188
Proposed method	0.7809	**0.9600**	**0.7269**	0.7065	0.8081	**0.9037**	0.9032	0.8700	0.9080	0.9793	**0.8546**



(a) Results of the *skating* sequence.



(a) Results of the *fence* sequence. (b) Results of the *woman* sequence.

❖ Video object segmentation measure

Measure	Mean $\mathcal{J}$	Recall $\mathcal{J}$	Mean $\mathcal{F}$	Recall $\mathcal{F}$
ViBe* [6]	0.2095	0.1364	0.1717	0.0773
FIC* [8]	0.1701	0.0607	0.2256	0.1337
BMRI-ViBe* [9]	0.1640	0.0553	0.1703	0.0817
MCD NP [25]	0.2634	0.0580	0.5569	0.7090
MCD 5.8ms [26]	0.3736	0.3756	0.5427	0.6100
Stochastic approx [28]	0.3398	0.3789	0.4003	0.4245
FP Sampling [27]	0.4294	0.5009	0.6031	0.7156
SC MCD [29]	0.5213	0.5952	0.7021	0.8200
uNLC [32]	0.1073	0.1002	0.1416	0.1181
OSVOS [1]	0.2547	0.2259	0.4129	0.3068
CIS [33]	0.1583	0.0591	0.2356	0.1253
BASNet [34]	0.5540	0.6204	0.6696	0.6880
Proposed method	**0.5603**	**0.6541**	**0.7214**	**0.8378**

$\mathcal{J}$: region-based segmentation similarity $\mathcal{F}$: contour-based accuracy

❖ Ablation study of two backgrounds

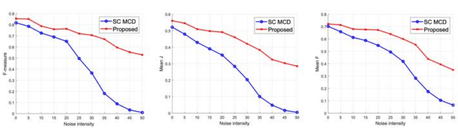
Method	*precision*	*recall*	F-measure
coarse BG model	0.9084	0.7655	0.8248
fine BG model	0.5669	0.7833	0.6095
combined model	0.9286	0.8041	0.8546

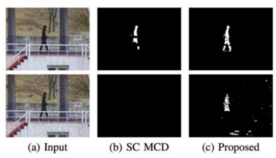
❖ Computational loads

CPU only, 320 x 240, 45.5fps

Module	Time (millisecond)
Motion estimation	2.207
Motion compensation	5.117
Age map update	0.595
Background model update	13.902
Foreground combining	0.671
Total	22.492

❖ Robustness test to noise image

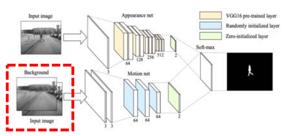




(a) Input (b) SC MCD (c) Proposed

Fig. 7. Example results for the noisy images. Each row shows the experimental results when the noise mean of image is 25 and 50, respectively.

❖ Combined with supervised method (AMNet)



Method	F-measure
AMNet [37] using MCD 5.8ms [26]	0.8789
AMNet [37] using SC MCD [29]	0.9175
AMNet [37] using Proposed BG	0.9529

References

NLC – A. Faktor and M. Irani, "Video Segmentation by Non-Local Consensus voting,," in *BMVC*, 2014.
OSVOS – S.Caelles et al., "One-Shot video object segmentation," in *CVPR*, 2017.
CIS – Y. Yang et al., "Unsupervised moving object detection via contextual information separation," in *CVPR*, 2019.
BASNet – X. Qin et al., "BASNet: Boundary-Aware Salient Object Detection," in *CVPR*, 2019.
ViBe – O. Barnich and M. Van Droogenbroeck, "ViBe: A Universal Background Subtraction Algorithm for Video Sequences." *IEEE Trans Image Process*, 2011.
FIC – J. Choi et al., "Robust moving object detection against fast illumination change," *Comput Vis Image Und*, 2012.
BMRI-Vibe – F. C. Cheng et al., "A background model re-initialization method based on sudden luminance change detection," *Eng Appl Artif Intell*, 2015.
MCD NP – Kim et al., "Detection of moving objects with a moving camera using non-panoramic background model," *Mach Vis Appl*, 2012.
MCD5.8ms – Yi et al., "Detection of Moving Objects with Non-stationary Cameras in 5.8ms: Bringing Motion Detection to Your Mobile Device," in *CVPR Workshop*, 2013.
Stochastic approx – F. J. Lo´pez-Rubio and E. Lo´pez-Rubio, "Foreground detection for moving cameras with stochastic approximation," *Pattern Recognit Lett*, 2015.
FP Sampling – K. Yun and J. Y. Choi, "Robust and fast moving object detection in a non-stationary camera via foreground probability based sampling." in *ICIP*, 2015.
SC MCD – K. Yun et al., "Scene conditional background update for moving object detection in a moving camera," *Pattern Recognit Lett*, 2017.
AMNet – Heo et al., "Appearance and motion based deep learning architecture for moving object detection in moving camera," in ICIP, 2017.