Appearance & Motion based Deep Architecture for Moving Object Detection in Moving Camera

Byeongho Heo, Kimin Yun, and Jin Young Choi
Introduction

- Moving object detection
Introduction

• Fixed camera

• Moving camera
Introduction

- Background-centric method

Input video

Compare

Background model

Moving object detection
Introduction

• Background for moving camera
  • Transform background model with homography $H$

![Diagram showing homography applied to video and background model between two time steps](image)
Introduction

- Background for moving camera
  - Update background by the image of t+1 frame

![](image1.png)
Introduction

- Background contamination

  - Background model in moving camera is not perfect
  - Motion compensation is not suitable for complex camera movements
  - Background based method is weak to background problem
Deep architecture

• Proposed method
  • Uses background model for motion information
  • Two components to cope with background contamination

1. Appearance information of moving objects
2. Learning based motion information
Deep architecture

- Structure of the proposed method
  - Input: image and background
  - Two sub-network: Appearance Net and Motion Net
  - Fully Convolutional Neural Network
Deep architecture

• Appearance network

  • A network without background
  • Detects appearance of movable objects
  • VGG-16 pre-trained network (objectness information)

**Appearance network**

Input image

- VGG16 pre-trained layer
- Zero-initialized layer

Soft-max
Deep architecture

• Motion network
  • Detects motion based on background image
  • Training dataset includes contaminated background
  • Randomly initialized shallow network (low-level information)

Motion network

Input image

Background

Randomly initialized layer

Zero-initialized layer

Soft-max
Deep architecture

- Merging appearance and motion
  - Unbalance between pre-trained and randomly initialized network
  - Two networks are separately trained for the balance
  - After that, two networks are merged and fine-tuned
Deep architecture

• Merging appearance and motion
  • Unbalance between pre-trained and randomly initialized network
  • Two networks are separately trained for the balance
  • After that, two networks are merged and fine-tuned
Deep architecture

- Merging appearance and motion
  - Unbalance between pre-trained and randomly initialized network
  - Two networks are separately trained for the balance
  - After that, two networks are merged and fine-tuned
Experiments

• Analysis

Daimler Results

Input video

Background

MCD in 5.8ms

Appearance Net

Motion Net

Proposed
Experiments

• Results

Video #1: Cycle

Input video

Proposed

MCD in 5.8ms[4]

FP sampling [6]

Stochastic approx [16]
Experiments

• Results

Video #2: Fence

Input video

Proposed

MCD in 5.8ms [4]

FP sampling [6]

Stochastic approx [16]
Experiments

• Results

Video #4: Campus

Input video

Proposed

MCD in 5.8ms[4]

FP sampling [6]

Stochastic approx [16]
Experiments

• Results

Video #5: Daimler

Input video

Proposed

MCD in 5.8ms[4]

FP sampling [6]

Stochastic approx [16]
Experiments

- **Performance**
  - Computation: 50 fps in GPU

MCD in 5.8 ms: “Detection of moving objects with nonstationary cameras in 5.8ms” in CVPR Workshops, 2013
FP sampling: “Robust and fast moving object detection in a non-stationary camera via foreground probability based sampling,” ICIP, 2015
We proposed a deep learning architecture that detects a moving object in a moving camera based on a background model.

To cope with background contamination, we designed the structure to use appearance information and learn the background pattern.

The proposed method detects moving objects robust to the background contamination and shows better performance than state-of-the-art methods.
Thank you
Experiments

• Details

• Training data : 14 videos
• Test data : 5 videos

• Input : 320 x 240 resolution
• GPU : GTX 1070

• Network : 14ms
• Background : 6ms
• Real-time 50fps