HENSOLDT Analytics

Generating Robust Motion Maps to Enhance Object Detector Performance

Sébastien Grand

Supervisor: Dr. Jonathan Kobold

06/02/2023

ModIA - Enseeiht

Table of contents

Introduction

- Context & Objectives
- Previous Work
- Requirements & Literature Review

Context & Objectives

Do you see any object on this frame ?

Figure 1 : Frame from Analytics drone dataset

Context & Objectives

<u>Would you have seen the drone on this frame if it did not have a bounding box?</u>

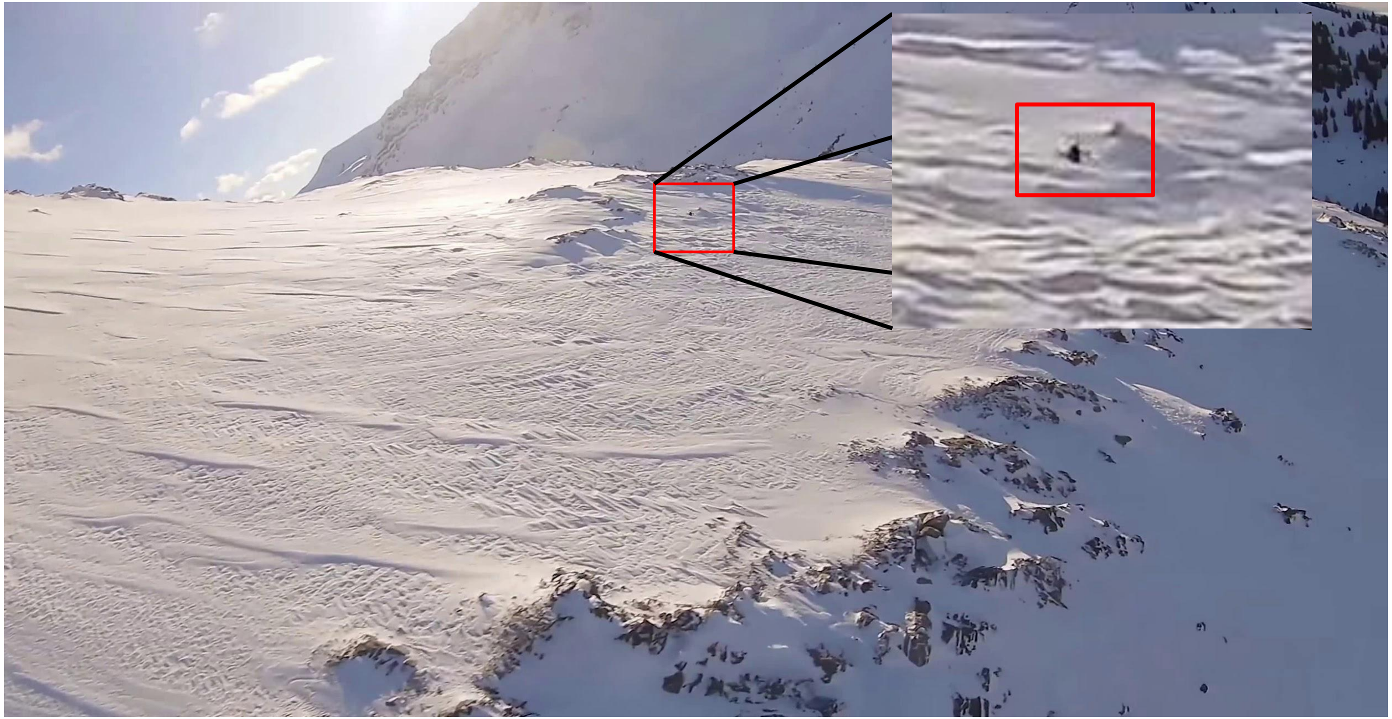


Figure 2 : Frame from Analytics drone dataset

Previous Work

For *visibility*, the video from the *Davis* dataset [1] called *BMX-bumps* is used as an example of *motion detection*.

Figure 3 : 10th frame from BMX-bumps video

Motion Detection Model

Figure 4 : Frame-difference computed with 2 frames

Figure 5 : Frame-difference computed with 3 frames [2]

Requirements & Literature Review

- <u>List of Requirements:</u>
  - The method must be executable on the *CPU* :
    - No research on *Deep Learning* methods.
  - The method should handle various situations :
    - Moving Camera;
    - Low-contrast;
    - Dynamic Background;
    - Tiny Objects

- <u>Literature Review:</u>
  - Optical Flow;
  - Panorama creation;
  - Subspace Learning;

Motion Detection : PRPCA

- Video Registration : Panorama Creation
- Motion Detection : Robust PCA Formulation
- Motion Detection : Decomposition Results
- Post-processing : Object Detection
- Results & Discussions

Video Registration : Frames 1

My camera captures a first frame.

Figure 6 : 1st frame from BMX-bumps video

Video Registration : Frames 2

My camera captures a second frame.

Figure 7 : 2nd frame from BMX-bumps video

Video Registration : Frames 3

My camera captures a third frame.

Figure 8 : 3rd frame from BMX-bumps video

Video Registration : Panorama of 3 frames

It is possible to align all the frames together with respect to a point of view reference.

Figure 9 : 3rd frame from BMX-bumps video

Video Registration : Panorama Creation

Video Registration procedure as proposed in [3].

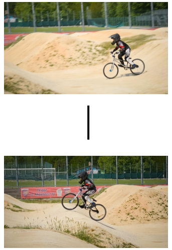


Figure 10 : Batch of N frames

Feature Extraction

Matching Features

Figure 11: Feature Extraction & Matching

Homographies Estimation

$$\begin{pmatrix} \mathcal{H}_{2 \to 1} \\ \cdots \\ \mathcal{H}_{anchor \to k} \\ \cdots \\ \mathcal{H}_{N-1 \to N} \end{pmatrix}$$

Frame Alignment

using the Homographies

Figure 12: Registered frames

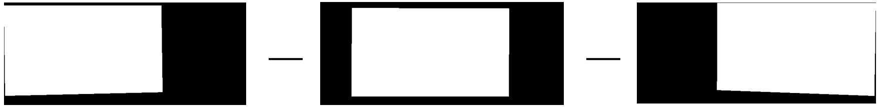


Figure 13: Support of the Registered frames

Video Registration : Resulting Panorama

With a batch of 20 frames captured by moving camera, it is possible to obtain this panorama :

Figure 14: Resulting panorama made of 20 frames
with the 10th frame as reference

Motion Detection : Separating the Motion from the Background

How can we separate what is moving (the bike) from what is static (the background) ?

a) Orignal RGB frame

b) Background

c) Motion Detected

Figure 15: Foreground/Background Segmentation

Motion Detection : Separating Motion from the Background

Problem formulation :

- Registered frames are stored in X as :

$$X = [vec(\tilde{x}_1), \ldots, vec(\tilde{x}_p)]$$

- where, vec(.) is the function to vectorize a frame;
- where $\tilde{x}_k$ refers to the *k*-th registered frame of *X* with respect to the *anchor frame*.
- Then, *X* can be decomposed [4] as :

$$\begin{aligned} \min_{L,S} \quad & ||L||_* + \lambda ||S||_1 \\ \text{subject to} \quad & \mathcal{P}_M(L+S) = \mathcal{P}_M(X) \end{aligned}$$

Motion Detection : Resulting Decomposition

Decompostion of *Paragliding* video from *Davis* dataset :

Figure 16: Paragliding Foreground/Backgroud Segmentation

Decompostion of drone video from Analytics dataset :

Figure 17: Foreground/Backgroud Segmentation of Drone video from Analytics Dataset

Post-processing : Object Detection Input Creation

Creation of *Binary Motion Map* & *Trajectory Map* :

Figure 18: RGB frame

Motion Detection with RPCA

Figure 19: Raw Motion Map

Otsu [5] Thresholding

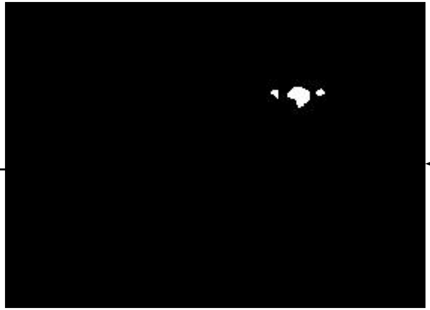


Figure 21: Binary Motion Map

Binary Motion Map
Accumulation

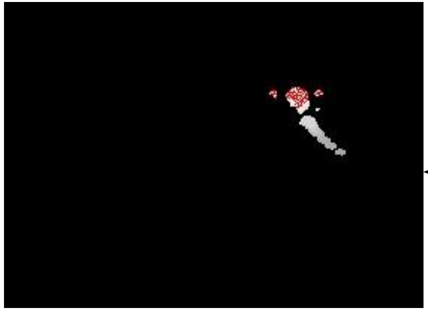


Figure 20: Trajectory Map

Results : Datasets & Metrics

- The two datasets used for *evaluation* are :
  - *Davis* Challenge dataset :
    - Dataset for *Background/Foreground Separation*;
    - The filming *camera is moving*;
  - *ChangeDetection.net* [6] dataset :
    - Database for testing *change detection* algorithm;
    - The filming camera jitters or the background is dynamic;

Figure 22:Motion Detection Result on BMX video

Compared to

Figure 23: Ground Truth of BMX-bumps segmentation

Results : Results & Discussions

Results of the experiments conducted on the Davis and CDnet datasets:

Video informations				PRPCA			Baseline
Dataset	Sequence	Camera	Frames	Precision	Recall	F-measure	F-measure
Davis 2017	bmx-bumps	moving	40	0.78	0.72	**0.75**	0.53
Davis 2017	tennis	moving	35	0.89	0.54	**0.67**	0.45
CDnet 2014	canoe	jittering	80	0.25	0.26	**0.22**	0.11
CDnet 2014	fall	jittering	80	0.80	0.53	**0.61**	0.22

Table 1: Results of PRPCA and baseline model on 4 different videos.

- True Positive (TP) : The model prediction matches the ground truth;
- False Positive (FP) : The model predicted a label that does not match the ground truth;
- False Negative (FN) : The model does not predict a label, but a ground truth label exists;
- Precision (P) : TP / (TP + FP)
- Recall (R) : TP / (TP + FN)
- F-measure/F1-score: 2 PR/ (P + R)

Object Detection

- Dataset for Object Detection
- Input Data for Object Detection
- Training Procedure
- Experiments & Results

Dataset for Object Detection

Slow Motion Camera Dataset

- Characteristics :
  - Easy motion compensation;
  - Tiny drones;
  - Low-contrast;
  - 13 videos;
  - ~50 frames / video;

Figure 24: Example of *RGB frame* and *Raw Motion Map* in *slow motion* dataset

Dataset for Object Detection

High Motion Camera Dataset

- Characteristics :
  - Difficult motion compensation;
  - More visible drones;
  - 11 videos;
  - ~50 frames / videos;

Figure 25: Example of *RGB frame* and *Raw Motion Map* in *high motion* dataset

Dataset for Object Detection

Slow & High Motion Camera Dataset

- Characteristics :
  - Various difficulty of motion compensation;
  - Various drone sizes;
  - Various image quality;

Input Data for Object Detection

The different *inputs* :

- *3-channel* input :
  - RGB frame (Baseline)
- *4-channel* input :
  - RGB frame + Raw Motion Map
  - RGB frame + Binary Motion Map
- *6-channel* input :
  - RGB frame + Trajectory Map

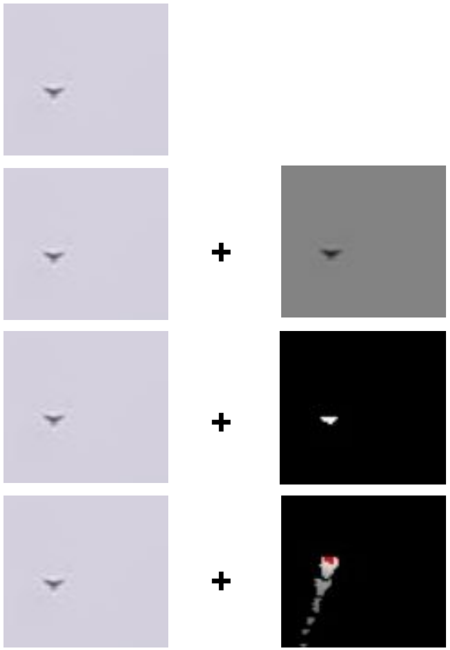
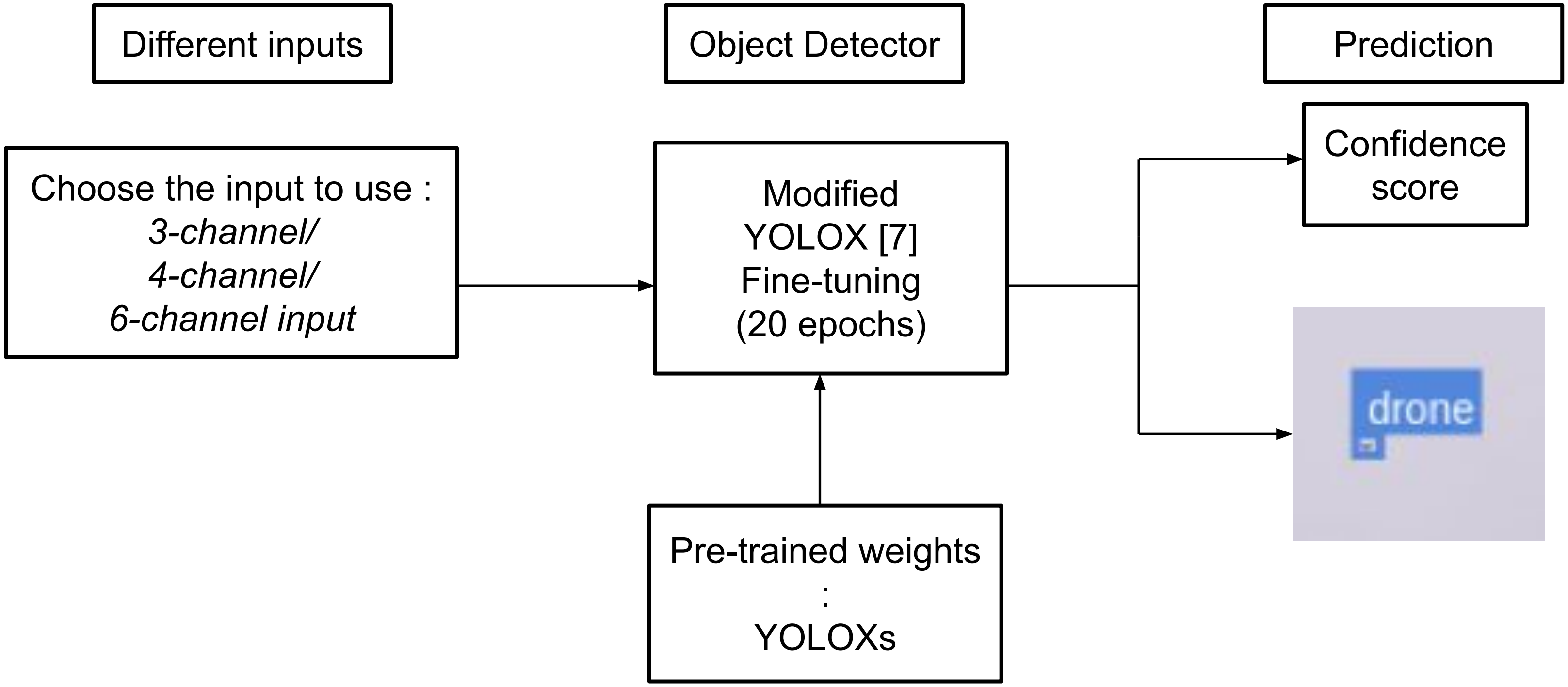


Figure 26: The differents inputs for object detection experiments

Training Procedure



Experiments & Results

Results of the experiments conducted on the Drone datasets of Analytics:

Dataset	Input	$AP_{drone}(\%)$	$AR_{drone}(\%)$
Slow motion camera	RGB	49.7	53.2
	RGB - Raw motion	52.3	54.2
	RGB - Binary motion	**58.4**	**62.0**
	RGB - Trajectory	54.5	57.0
High motion camera	RGB	**56.4**	**61.0**
	RGB - Raw motion	55.7	58.0
	RGB - Binary motion	48.1	52.1
	RGB - Trajectory	52.2	54.0
Coupled dataset	RGB	45.6	49.8
	RGB - Raw motion	51.2	55.7
	RGB - Binary motion	49.7	52.0
	RGB - Trajectory	**53.7**	**56.3**

Table 2: YOLOX results on different datasets (Analytics)

AP (%) : average of all precision measurements for each detection.
AR (%) : average of all recall measurements for each detection.

Figure 27: Accurate Motion Map (good ego -motion compensation

Figure 28: Noisy Motion Map due to high ego-motion

Possible Improvements & Future Work

- Future improvements :
  - Motion compensation :
    - Extract better features;
    - Estimate better Homographies;
  - Elimination of clutter (moving trees, flowing water):
    - CNN to train to remove clutter;

Conclusion

- In this work, I was able to :
- Learn and perform *motion compensation* by panorama building approach;
- Learn and perform *motion detection* with *Robust PCA;*
- *Improve motion detection* compared to previous work;
- Fine-tune an object detector with different inputs;
- Show that adding *relevant motion information* to an RGB image as input to an object detector *improves* its performance at detecting *small drones*.

Thanks for your attention.

Any questions ?

References

[1] DAVIS: Densely annotated VIdeo segmentation
[2] G.-W. Yuan, J. Gong, M.-N. Deng, H. Zhou, and D. Xu, “A moving objects detection algorithm
based on three-frame difference and sparse optical flow,“
[3] B. E. Moore, C. Gao, and R. R. Nadakuditi, “Panoramic robust PCA for foreground-background separation on noisy, free-motion camera video.”
[4] E. J. Candes, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?”
[5] N. Otsu, “A threshold selection method from gray-level histograms,”
[6] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, “CDnet 2014: An expanded change detection benchmark dataset,”
[7] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, “YOLOX: Exceeding YOLO series in 2021.”