Background subtraction and background generation

Marc Van Droogenbroeck

Department of Electrical Engineering and Computer Science
University of Liège

August 2016

What is background subtraction in computer vision?

There are basically two "pure" approaches for motion analysis in a video sequence:

1. Motion analysis by *tracking* (= *motion estimation* based techniques):
    - detects some particular points (*features*) in a video frame.
    - find the corresponding points in the next frame.
    - based on a model, interpret the trajectories (called *tracks*) of the points (usually at the object level).
2. Motion analysis by *background subtraction*:
    - build a reference frame or model with no foreground in it.
    - compare a next frame to the reference.
    - extract foreground objects.

Number of papers on background subtraction

Searches in databases with the following keywords:

background subtraction video

Databases with pdf documents	2015	$\longrightarrow$ today
IEEE Explore	137	1,272
Springer Link	1,207	8,460
Elsevier (ScienceDirect)	774	7,169
google scholar	16,300	144,000

Motion analysis by background subtraction I

Original image

Segmentation map

- This is a classification problem (with two classes) that separates the *foreground* (pixels "in motion") from the *background* ("static" pixels).
- Evaluation via classification notions such as the *precision*, *recall*, *ROC space*, F_1 score, *error rate*, etc.

Steps in background subtraction

[Initialization] build a *reference frame* or a *statistical model* for the background.

[Subtraction or segmentation] *compare* the current frame to the reference frame or model, and "*subtract*" the frame to get a binary image indicating pixels who have changed.

[Updating] *update* the reference frame or model.

When developing a technique, we have to detail these three steps! This is why there are so many variants.

Rough typology of methods

1. Estimation of the probability distribution function (pdf) for each pixel location ($\Rightarrow$ statistical models)
   - Mixture of Gaussians MoG (parametric methods): estimate the mean + standard deviation
   - **Kernel Density Estimation KDE (non-parametric methods): estimate the pdf from past samples**
2. Techniques based on learning/dictionaries
   - Codebooks
   - Bag of words
3. Techniques based on data reduction
   - Robust PCA

Implementation issues

A video sequence is like a data *cube* whose dimension is only fixed in 2 (spatial) dimensions.

- Cube extends with time.
- Although the use of "memory" should be kept constant.

Challenges:

1. need to find a way to accumulate knowledge of increasing size inside of a constant-sized memory block.
2. this knowledge should be updated regularly to deal with changes (to understand the challenge, think of a camera operating day and night).

Background generation

Definition (Background generation)

Given a scene viewed from a fixed viewpoint, the problem of generating an image of the background is known as the *background generation problem*.

Figure: The generation of a stationary background image is a challenging task, especially when the background is never fully visible.

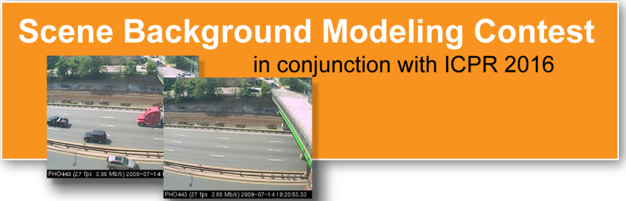
Comparison between background subtraction and background generation

background subtraction	**background generation**
infinite data volume	finite data volume
should be universal	scene specific
evaluation via classification metrics: *F_1 score* (↑)	evaluation via PSNR, error rate: *percentage of Error Pixels* (pEPs, ↓)

Background generation is closer to what we do in stellar imaging.

IEEE Scene Background Modeling Contest (July 2016)

http://pione.dinf.usherbrooke.ca/sbmc2016/ : 79 videos out of 8 categories



Results I

Method	Average ranking across categories	Average ranking
LaBGen [6]	4.25	3.33
LaBGen-P [7]	4.88	4.50
Temporal median filter [2]	5.13	6.67
SC-SOBS-C4 [9]	5.63	4.67
Bidirectional Analysis and Consensus Voting [12]	5.75	7.33
Bidirectional Analysis [28]	5.75	6.67
Wei-Liu-Aug-16-2 [11]	5.88	8.33

[6] LaBGen: A Method Based on Motion Detection for Generating the Background of a Scene, B. Laugraud, S. Piérard and M. Van Droogenbroeck, to appear in *Pattern Recognition Letters*, 2016.

Results II

LaBGen mechanisms

LaBGen uses:

1. a temporal median filter.
2. a pixel/patch based motion detection algorithm (via *background subtraction* techniques).
3. + other minor refinements.

Results III

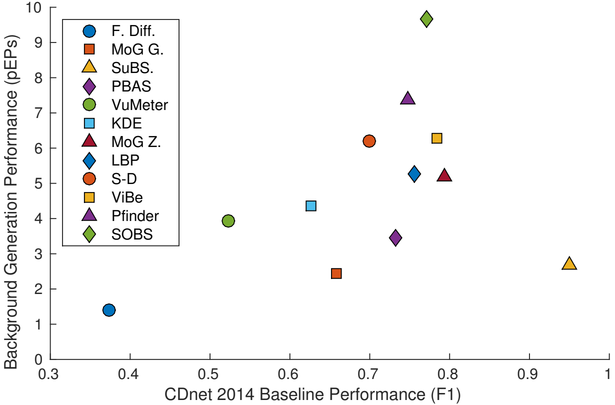


Figure: Comparison between the F_1 (↑) performance of a background subtraction algorithm on the baseline of the CDnet 2014 dataset and the pEPs score (↓).

Results IV

Conclusions

- The quality of *background generation* **is unrelated to** the *background subtraction* method: no (negative) correlation.
- The temporal median filter is within the top techniques (remember ADI/LOCI)

What's next? I

More machine learning (not yet for background generation?!)

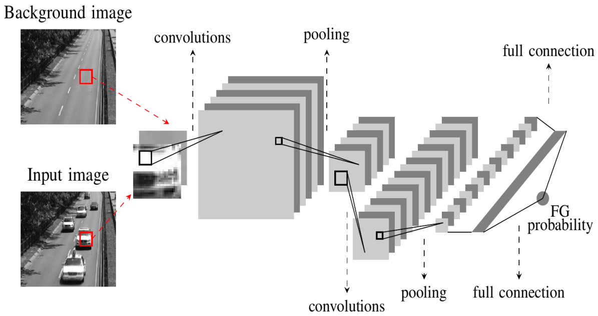


Figure: Deep learning for extracting the background in video scenes (M. Braham and M. Van Droogenbroeck. **Deep Background Subtraction with Scene-Specific Convolutional Neural Networks**. In *IEEE IWSSIP*, May 2016).

What's next? II

Method	$F_{overall}$	$F_{Baseline}$	F_{Jitter}	$F_{Shadows}$	$F_{LowFramerate}$
ConvNet-GT	**0.9046**	**0.9813**	**0.9020**	**0.9454**	**0.9612**
IUTIS-5	0.8093	0.9683	0.8022	0.8807	0.8515
SuBSENSE	0.8018	0.9603	0.7675	0.8732	0.8441
PAWCS	0.7984	0.9500	0.8473	0.8750	0.8988
PSP-MRF	0.7927	0.9566	0.7690	0.8735	0.8109
ConvNet-IUTIS	0.7897	0.9647	0.8013	0.8590	0.8273
EFIC	0.7883	0.9231	0.8050	0.8270	0.9336
Spectral-360	0.7867	0.9477	0.7511	0.7156	0.8797
SC_SOBS	0.7450	0.9491	0.7073	0.8602	0.7985
GMM	0.7444	0.9478	0.6103	0.8396	0.8182
GraphCut	0.7394	0.9304	0.5183	0.7543	0.8208

Table: Overall and per-category F scores for different methods.