An Overview of Background Initialization and LaBGen

Benjamin Laugraud

Montefiore Institute, University of Liège, Belgium

April 3rd, 2018

IEEE International Conference on Intelligent Systems and Computer Vision (ISCV 2018)

Fez, Morocco

Introduction to Background Initialization

What is Background Initialization?

Definition

Given a video sequence acquired from a static viewpoint, the ***stationary background initialization*** problem (also known as *background generation*, *estimation*, *extraction*, or *reconstruction* problem) consists in **generating a unique image estimating the stationary background of the sequence** (*i.e.* the set of elements which are motionless throughout the sequence). (Laugraud et al. 2016)

Example: Recovering a Road Without Cars



↓

What is Background Initialization?

- **Several applications** $\rightarrow$ video surveillance, computational photography, etc (see Maddalena et al. 2015).

- Not as easy as it looks $\rightarrow$ there are **several challenges**!

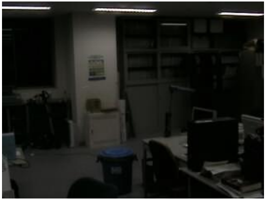
- Let's see the **most important challenges** described in Jodoin et al. 2017.

- Note that the images used for illustrating the challenges have been taken from a **public dataset (SBMnet)**.

Illumination Changes Challenge

- **Light or strong** illumination changes (*e.g.* light switching, weather, etc).
- **Background evolves** over time → several solutions.
- Methods should pay attention to **temporal order**.

BACV
(Minematsu et al. 2016)



Expected output 1

Expected output 2

Clutter Challenge

- Foreground objects occupying a **large portion of the visual field**.
- Several pixels depict foreground objects **more than 50% of the time**.
- A **pixel-wise temporal median filter** cannot cope with this kind of sequences.

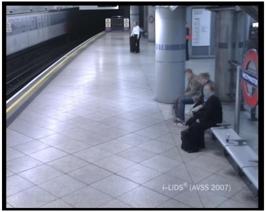
Temporal median

Input frame

Expected output

Intermittent Motion Challenge

- Objects that **stop for a short while**.
- Objects that are **abandonned**.
- Background objects **starting to move**.



Temporal median

LaBGen
(Laugraud et al. 2017b)

Expected output

- **Unstable** camera (*e.g.* wind, vibrations in surrounding environment).
- Consequently, the background is **also in motion**.
- Methods should discover and **compensate for the camera motion**.

RMR
(Ortego et al. 2016)

RSL2011
(Reddy et al. 2011)

Expected output

- Video sequences with a **limited number of frames**.
- Very **low** frame rate, or **no temporal** order.
- Increases the **difficulty of detecting motion** for traditional models.

Temporal median

BE-AAPSA
(Ramirez-Alonso et al. 2017)

Expected output

Brief Overview of Some Popular Methods

(Following the Taxonomy of Bouwmans et al. 2017)

Properties

- Based on statistics (*e.g.* mean, median) computed on **temporal information**.
- Statistics computed pixel-wise on the **whole sequence or random frames**.

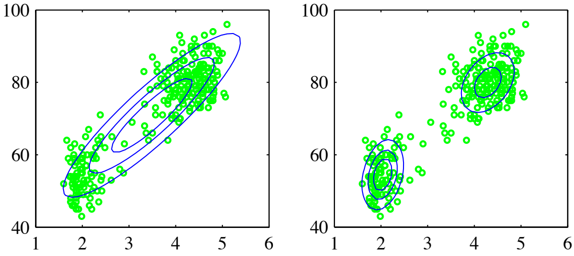
Temporal Median

- Pixel-wise **temporal median** filter considering **all frames**.
- Assumption → the background is observed **> 50% of the time** in each pixel.
- This assumption is false in highly **cluttered sequences**.
- But it produces **excellent** results for **basic scenarios** (see SBMnet basic cat.).

Mixture of Gaussians (Stauffer et al. 1999)

- **Background subtraction** technique (in part).
- Background **model** $\rightarrow$ mixture of K **Gaussians** per-pixel.
- **Proba. of observing** current pixel value $\rightarrow$ determined by **associated mixture**.
- Distributions **adapted over time** using an **online K-means** approximation.
- **Background image** generation $\rightarrow$ **weighted average** of background modes.

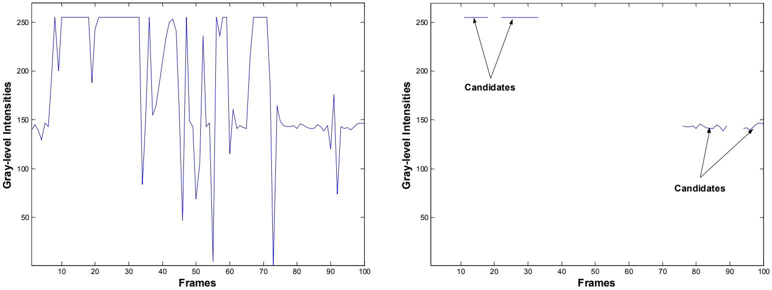
(Bishop 2006)



Properties

- **Assumption:** background has the **longest stable intensity**.
- Stable temporal subsequences are **located**, and the most reliable is **chosen**.

(Wang et al. 2006)



WS2006 (Wang et al. 2006)

1. **Locate** non-overlapping stable subsequences of pixel intensities (SSIs):
   - Find all subsequences meeting **three criteria**.
   - The subsequence has a **minimum size** L_w.
   - Difference between all temporally **consecutive intensities** $< T_f$ (a threshold).
   - Difference between an **intensity** and the **mean of all previous intensities** $< T_f$.
2. **Choose** the most reliable SSI:
   - Compute **size and variance** of each selected subsequence.
   - Keep the one **maximizing the ratio** between both.

Methods Based on Iterative Model Completion

Properties

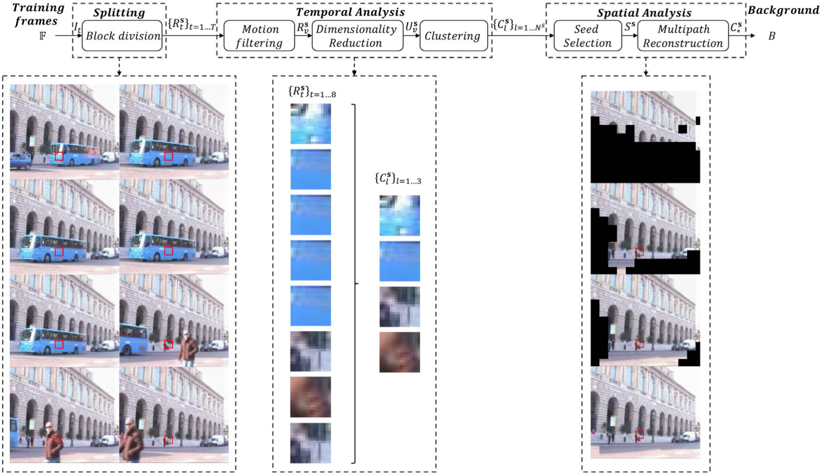
- **Spatial areas** with a **static reliable background** are generated.
- The **remaining areas** are **completed** according to **spatial consistency criteria**.

RMR (Ortego et al. 2016)

- Each frame is divided in **s** patches $R_t^{\mathbf{s}}$ at time t.
- In each spatial area **s** $\rightarrow$ ***motion filtering*** to discard patches in motion (f. diff.).
- ***Custering*** performed to build a set of candidates $C_l^{\mathbf{s}}$.
- **Several numbers of clusters** $N^{\mathbf{s}}$ tested $\rightarrow$ **between 1 and the number of SSIs** detected during motion filtering.
- **Choose** $N^{\mathbf{s}}$ $\rightarrow$ **metric** max. compactess and separation, and min. similarity.
- **Candidates** $C_l^{\mathbf{s}}$ in spatial area **s** $\rightarrow$ **mean** of each cluster $K_l^{\mathbf{s}}$, with $l = 1, \ldots, N^{\mathbf{s}}$.
- ***Seed selection*** selects **highly reliable candidates** $S^{\mathbf{s}}$ with a large cluster cardinality and a low motion activity in the associated spatial area.
- Spatial areas **s** with **empty seeds** $S^{\mathbf{s}} = \emptyset$ are **iteratively completed** $\rightarrow$ inter- and intra-block smoothness constraints (connected neighborhood).

Methods Based on Iterative Model Completion

(Ortego et al. 2016)



Methods Based on Optimal Labeling

Properties

- Find a **label** in each pixel/region indicating the **frame n° with background**.
- Consists in minimizing a **spatio-temporal cost function**.

Photomontage (Agarwala et al. 2004)

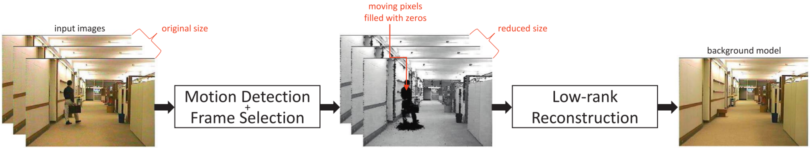
- Create a ***composite*** image from a ***stack of source*** images.
- To the composite is associated a ***labeling*** specifying the source for each pixel.
- A **pixel labeling** is determined through **graph-cut optimization**.
- The ***cost function*** is defined as the sum of an ***image and seam objective***.
- The **image objective** is the **maximum likelihood** (the most common value).
- Probability distribution $\rightarrow$ **color histogram** of the same pixel in the sources.
- The result is refined with **gradient-domain fusion**.
- A **vector field** is build from the **optimal labeling**, then the fusion is applied.

Methods Based on Missing Data Reconstruction

Properties

- Pixels/regions in **motion** are considered as **missing data**.
- Missing **data is recovered** using inpainting or low-rank reconstruction.

Matrix Completion (Sobral et al. 2017)



- Reconstruction has a high computational cost → **remove redundant frames**.
- A vector **d** containing the L_2 **distances** between **consecutive frames** is built.
- From its derivative **d**′, **non-redundant** frames are **selected**.
- The selected frames are **vectorized** and put in a **matrix A**.
- The values of **A** far from their temporal predecessor (**motion**) are set to **zero**.
- A **matrix completion** algorithm is applied on **A**.

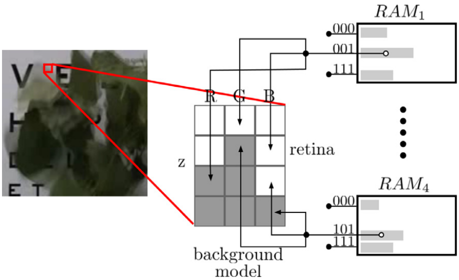
Properties

- **Learn** automatically the background **from the data**.
- The learning can be **supervised** or **unsupervised**.

BEWiS (De Gregorio et al. 2017)

- **Pixel-level** method based on the WiSARD^{rp} **weightless neural network**.
- Learning from the sequence without any annotation $\rightarrow$ **unsupervised**.

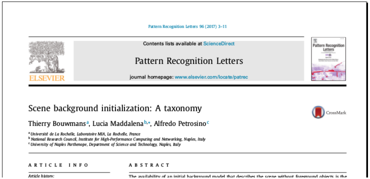
(De Gregorio et al. 2017)



- Channels of a pixel value are first **scaled** between $[0, \ldots, z-1]$ and **binarized**.
- The result forms a **binary pattern** observed by a *retina*.
- The bits of the retina are **randomly mapped** to a set of *n*-**tuple RAM neurons**.
- For each RAM, the cell at the address read from the retina $\rightarrow$ **reward** ρ.
- Other cells $\rightarrow$ decreased by a **punishment** Ψ.
- Background recovered from the **addresses** of the cells with the **highest values**.

To Get Further Information

There are two major reviews!


Pattern Recognition Letters 96 (2017) 3–11

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

ELSEVIER

Scene background initialization: A taxonomy

Thierry Bouwmans[a], Lucia Maddalena[b,*], Alfredo Petrosino[c]

[a] Université de La Rochelle, Laboratoire MIA, La Rochelle, France
[b] National Research Council, Institute for High-Performance Computing and Networking, Naples, Italy
[c] University of Naples Parthenope, Department of Science and Technology, Naples, Italy

ARTICLE INFO

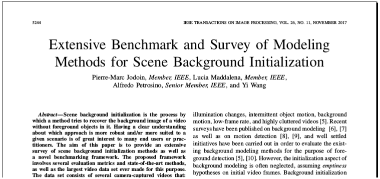
Article history:

ABSTRACT

The availability of an initial background model that describes the scene without foreground objects is the

(Bouwmans et al. 2017)

URL: https://doi.org/10.1016/j.patrec.2016.12.024


5244 IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 26, NO. 11, NOVEMBER 2017

Extensive Benchmark and Survey of Modeling Methods for Scene Background Initialization

Pierre-Marc Jodoin, *Member, IEEE*, Lucia Maddalena, *Member, IEEE*, Alfredo Petrosino, *Senior Member, IEEE*, and Yi Wang

***Abstract*—Scene background initialization is the process by which a method tries to recover the background image of a video without foreground objects in it. Having a clear understanding about which approach is more robust and/or more suited to a given scenario is of great interest to many end users or practitioners. The aim of this paper is to provide an extensive survey of scene background initialization methods as well as a novel benchmarking framework. The proposed framework involves several evaluation metrics and state-of-the-art methods, as well as the largest video data set ever made for this purpose. The data set consists of several camera-captured videos that:**

illumination changes, intermittent object motion, background motion, low-frame rate, and highly cluttered videos [5]. Recent surveys have been published on background modeling [6], [7] as well as on motion detection [8], [9], and well settled initiatives have been carried out in order to evaluate the existing background modeling methods for the purpose of foreground detection [5], [10]. However, the initialization aspect of background modeling is often neglected, assuming *emptiness* hypotheses on initial video frames. Background initialization

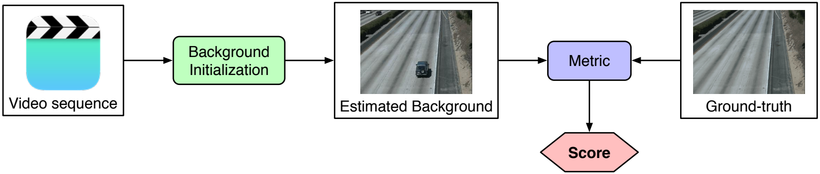
(Jodoin et al. 2017)

URL: https://doi.org/10.1109/TIP.2017.2728181

Evaluation Framework

Estimation vs. Ground-truth

- How to **assess and compare** methods? What about their **robustness**?
- **Low-level operation** → comparing an estimation to a **ground-truth image**.
- A ***ground-truth*** image depicts the **"perfect" background**.
- No unique solution → compare to **several ground-truths** (*e.g.* traffic light).
- **Difference** between an estimation and a ground-truth **quantified by a metric**.



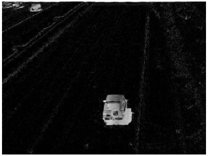
- Comparison to **several ground-truths** → keep the **best score**.

Metrics

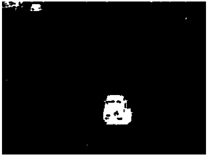
- **Average Gray-level Error (AGE):** mean absolute difference of luminance between estimation and ground-truth.
- **Percentage of Error Pixels (pEPs):** percentage of pixels whose absolute difference of luminance between estimation and ground-truth > 20.
- **Percentage of Clustered Error Pixels (pCEPs):** percentage of error pixels whose 4-connected neighbors are also errors.


Estimation


Ground-truth


Abs. Difference


EPs


CEPs

- **Multi-Scale Structural Similarity Index (MS-SSIM; Wang et al. 2003):** pyramidal SSIM, which uses structural distortion as an estimate of the perceived visual distortion. (Wang et al. 2004)

- **Peak Signal-to-Noise Ratio (PSNR):**

$$\mathrm{PSNR} = 10 \times \log_{10} \frac{255^2}{\mathrm{MSE}},$$

with MSE being the mean squared error.

- **Color image Quality Measure (CQM; Yalman et al. 2013):** combination of per-channel PSNRs computed on an approximated reversible RGB to YUV transformation.

- Using those metrics has been proposed by Maddalena et al. 2015.

Datasets

- On **which data** the metrics can be used?
- A ***dataset*** provides video sequences along with their ground-truth.
- **Public datasets** enable to perform **fair comparisons** in the literature.
- When several methods are evaluated on the same dataset $\rightarrow$ **ranking**.
- State-of-the-art datasets gather **video sequences in categories** (challenges).
- Enable to **evaluate their robustness** against different challenges.

Scene Background Initialization (SBI)

- Composed of **14 video sequences**.
- `http://sbmi2015.na.icar.cnr.it/SBIdataset.html`

Advantages

- **Ground-truth** provided for **all sequences**.
- Provides **Matlab scripts** for performance evaluation.

Drawbacks

- **Small number of video sequences** preventing a good generalization.
- The sequences are **not gathered by categories**.
- **No submission platform** for a public comparison of different methods.

SceneBackgroundModeling.NET (SBMnet)

- Composed of **79 video sequences**.
- Scattered through **8 categories**: Basic, Intermittent Motion, Clutter, Jitter, Illumination Changes, Background Motion, Very Long, and Very Short.
- `http://www.scenebackgroundmodeling.net` (Jodoin et al. 2017)

Advantages

- Uses **several ground-truths** when needed.
- **Various challenges** covered through the proposed categories.
- Provides **Python and Matlab scripts** for performance evaluation.
- Online **submission platform** allowing an up-to-date comparison of methods.

Drawbacks

- **Small number** of **ground-truths** provided (for 13 sequences only).
- **Two rankings** are proposed, and they sometimes disagree.

SBMnet: Average Ranking

HOME | RESULTS | DATASET | UTILITIES | UPLOAD | SBMC2016

Results for SBMnet 2016

Overall | Basic | Intermittent Motion | Clutter | Jitter | Illumination Changes | Background Motion | Very Long | Very Short

Results, all categories combined.

Click on method name for more details.

Method	Average ranking	Average ranking across categories	Average AGE	Average pEPs	Average pCEPS	Average MSSSIM	Average PSNR	CQM
MSCL [15]	1.17	4.88	5.9547	0.0524	0.0171	0.9410	30.8952	31.7049
LaBGen-OF [25]	2.17	4.50	6.1897	0.0566	0.0232	0.9412	29.8957	30.7006
BEWiS [24]	4.33	5.88	6.7094	0.0592	0.0266	0.9282	28.7728	29.6342
LaBGen [6]	4.83	7.75	6.7090	0.0631	0.0265	0.9266	28.6396	29.4668
NExBI [26]	6.00	11.38	6.7778	0.0671	0.0227	0.9196	27.9944	28.8810
LaBGen-P [7]	6.33	8.50	7.0738	0.0706	0.0319	0.9278	28.4660	29.3196
Photomontage [3]	7.17	11.13	7.1950	0.0686	0.0257	0.9189	28.0113	28.8719
SC-SOBS-C4 [9]	8.33	9.63	7.5183	0.0711	0.0242	0.9160	27.6533	28.5601
MAGRPCA [10]	9.50	9.88	8.3132	0.0994	0.0567	0.9401	28.4556	29.3152
Temporal median filter [2]	11.67	9.13	8.2761	0.0984	0.0546	0.9130	27.5364	28.4434

SBMnet: Category Average Ranking (Clutter)

HOME | RESULTS | DATASET | UTILITIES | UPLOAD | SBMC2016

Results for SBMnet 2016

Overall | Basic | Intermittent Motion | Clutter | Jitter | Illumination Changes | Background Motion | Very Long | Very Short

Results, for the clutter category.

Click on method name for more details.

Method	Average ranking	Average AGE	Average pEPs	Average pCEPS	Average MSSSIM	Average PSNR	CQM
LaBGen-OF [25]	1.17	4.1821	0.0246	0.0117	0.9640	32.6339	33.4654
MSCL [15]	1.83	5.2695	0.0275	0.0094	0.9629	31.3743	32.2837
NExBI [26]	3.33	5.3091	0.0414	0.0141	0.9379	30.0056	30.9847
Photomontage [3]	6.17	6.8195	0.0543	0.0294	0.8892	28.5554	29.4882
Bidirectional Analysis [13]	6.67	6.6565	0.0497	0.0177	0.9243	26.4376	27.5267
SC-SOBS-C4 [9]	7.33	7.0590	0.0644	0.0304	0.8939	28.0077	29.0737
MAGRPCA [10]	7.67	8.1589	0.0647	0.0294	0.9446	26.6872	27.5988
RMAMR [20]	7.67	8.1589	0.0647	0.0294	0.9446	26.6872	27.5988
RSL2011 [4]	8.33	7.3013	0.0701	0.0375	0.9087	27.9304	28.9763
LaBGen-P [7]	8.33	7.8947	0.0986	0.0678	0.8967	28.1140	29.1305

SBMnet: Method Details (LaBGen)

HOME RESULTS DATASET UTILITIES UPLOAD SBMC2016

Contact name	Benjamin Laugraud
Contact email	blaugraud@ulg.ac.be
Contact university/company	University of Liège
Method's name	LaBGen
Reference	B. Laugraud, S. Piérard, M. Van Droogenbroeck "LaBGen-P: A Pixel-Level Stationary Background Generation Method Based on LaBGen", Scene Background Modeling workshop (ICPR) 2016
Processing time	~1312 FPS for a 640x480 video with C++ code on a Core i7-4790K
Code is available online	True
Web page	http://www.telecom.ulg.ac.be/labgen
Parameters	A = frame_difference, P = 1 , N = 3 , S = 19

You can download the background estimation results of this method by clicking here.

Basic | Background Motion | Jitter | Intermittent Motion | Clutter | Illumination Changes | Very Long | Very Short

Video	AGE	pEPs	pCEPS	MSSSIM	PSNR	CQM
Blurred	1.3990	0.0001	0.0000	0.9975	41.5779	41.6541
PETS2006	2.0432	0.0027	0.0019	0.9896	35.2304	35.9487
Intersection	2.4286	0.0009	0.0000	0.9900	37.2576	37.7342

LaBGen

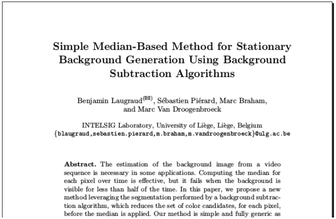
Ideas

1. Median buffer **polluted with foreground** elements $\rightarrow$ let's filter them out.
2. Such a filtering could be **motion-based**.
3. Discard elements with the **largest "quantities of motion"**.

Observation

Background subtraction (BGS) algorithms are designed to detect motion!

LaBGen in Short



Simple Median-Based Method for Stationary Background Generation Using Background Subtraction Algorithms

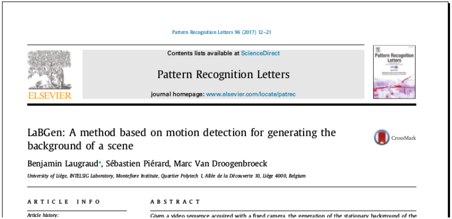
Benjamin Laugraud(✉), Sébastien Piérard, Marc Braham, and Marc Van Droogenbroeck

INTELSIG Laboratory, University of Liège, Liège, Belgium
{blaugraud,sebastien.pierard,m.braham,m.vandroogenbroeck}@ulg.ac.be

Abstract. The estimation of the background image from a video sequence is necessary in some applications. Computing the median for each pixel over time is effective, but it fails when the background is visible for less than half of the time. In this paper, we propose a new method leveraging the segmentation performed by a background subtraction algorithm, which reduces the set of color candidates, for each pixel, before the median is applied. Our method is simple and fully generic as

(Laugraud et al. 2015)

URL: http://hdl.handle.net/2268/182893



Pattern Recognition Letters 96 (2017) 12–21

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

ELSEVIER

LaBGen: A method based on motion detection for generating the background of a scene

Benjamin Laugraud*, Sébastien Piérard, Marc Van Droogenbroeck

University of Liège, INTELSIG Laboratory, Montefiore Institute, Quartier Polytech 1, Allée de la Découverte 10, Liège 4000, Belgium

ARTICLE INFO

Article history:

ABSTRACT

Given a video sequence acquired with a fixed camera, the generation of the stationary background of the

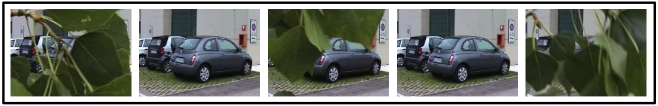
(Laugraud et al. 2017b)

URL: http://hdl.handle.net/2268/203572

- It combines a **pixel-wise median** filter and a **patch selection** mechanism.
- The selection mechanism is **based on motion detection** (BGS).
- This mechanism selects the **patches with the smallest amounts of motion**.
- The **pipeline** of the method comprises **5 steps**.

Step 1: Augmentation

- BGS needs sometimes a **long training** $\rightarrow$ problem with **short sequences**.
- The agumentation step **increases the duration** of the input video sequence.
- In practice, we process the sequence in $\mathcal{P}$ **passes**.
- An **odd** (resp. **even**) pass is performed **forwards** (resp. **backwards**).



forwards (odd passes)

backwards (even passes)

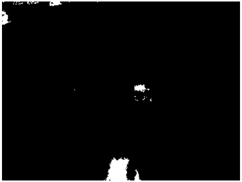
- ViBe (Barnich et al. 2011) is highly **sensitive to bootstrap**.
- First frame contains several foreground objects $\rightarrow$ **ghosts**!
- The **augmentation** mechanism can help to **reduce false alarms**.

(Laugraud et al. 2017b)

Frame 1

Frame 446

Pass 1



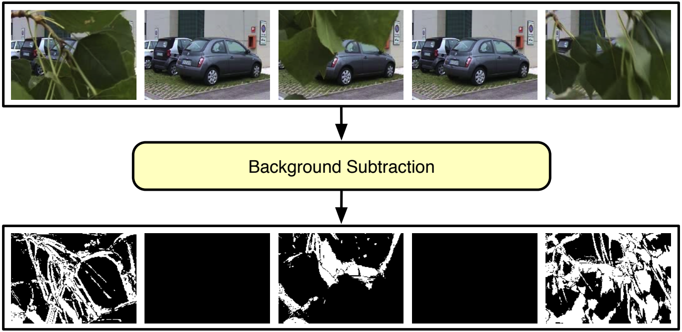
Pass 2

Pass 3

Pass 5

Step 2: Motion Detection

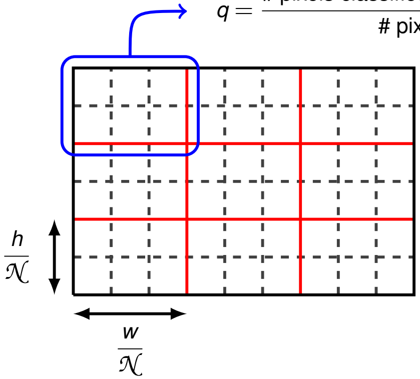
- The **BGS algorithm** being used is the **parameter** $\mathscr{A}$.
- LaBGen does not leverage the model of $\mathscr{A}$, **only segmentation maps**.
- LaBGen can be used with **any BGS algorithm** "out-of-the-box".



Step 3: Local Estimation of the Quantities of Motion

- The image plane is divided into $\mathcal{N} \times \mathcal{N}$ non-overlapping **spatial areas**.
- A **quantity of motion** q is estimated for each patch of each frame.
- **Probability** of observing pixels corresponding to **moving objects**.

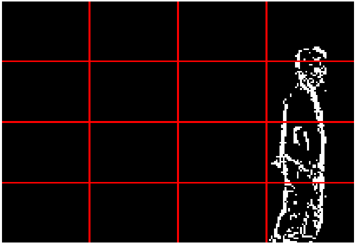
$$q = \frac{\text{\# pixels classified as foreground in the patch}}{\text{\# pixels in the patch}}$$



Why it is Necessary to Work at the Region Level?

- Some BGS algorithms are optimized to **reduce false positives**.
- Consequently, **false negatives** could be **increased**.
- No guarantee that more than 50% of pixels classified as background belong to the background (**negative predictive value** > 0.5).
- We can taking into account classifications in the **spatial neighborhood**.

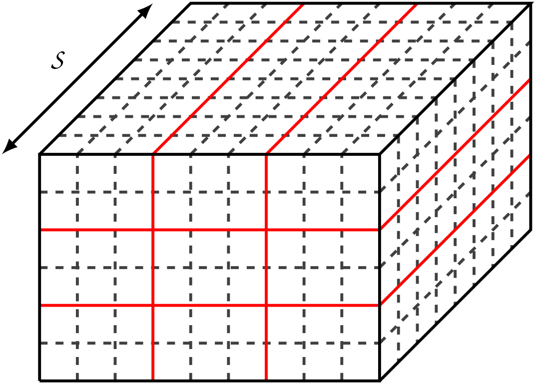
CaVignal - Frame 244



Frame Difference

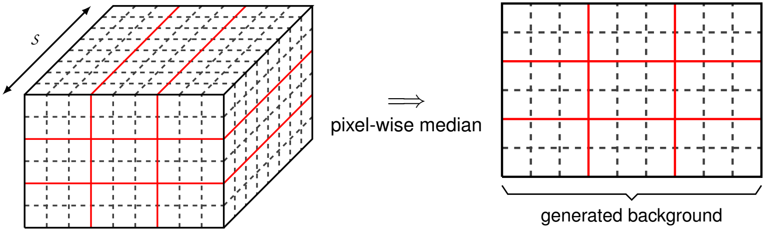
Step 4: Patch Selection

- In each spatial area, $\mathcal{S}$ **patches are selected**.
- The $\mathcal{S}$ selected patches are associated to the **smallest quantities of motion** q.

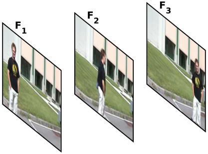


Step 5: Background Generation

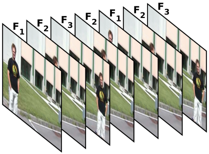
- A **pixel-wise median filter** is applied on the sets of $\mathcal{S}$ selected patches.
- The **background** is then **generated**.



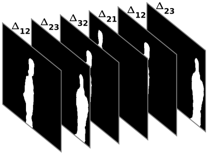
(Laugraud et al. 2017a)



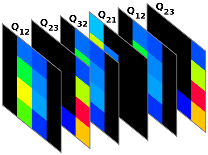
Input video sequence



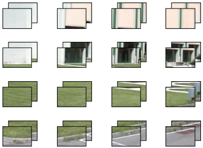
1. Augmentation step



2. Motion detection step



3. Estimation step

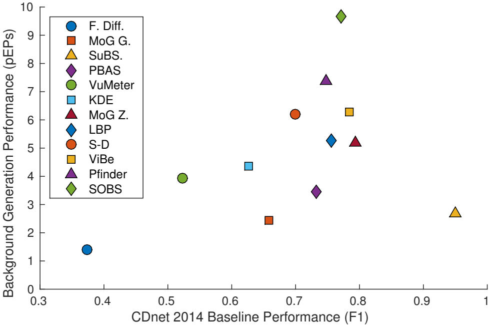


4. Selection step

5. Generation step (output)

- **No obvious correlation** between the performance of LaBGen and BGS.
- **Worst** (resp. best) BGS algorithm → 1st (resp. 3rd) **best** for LaBGen.

(Laugraud et al. 2017b)



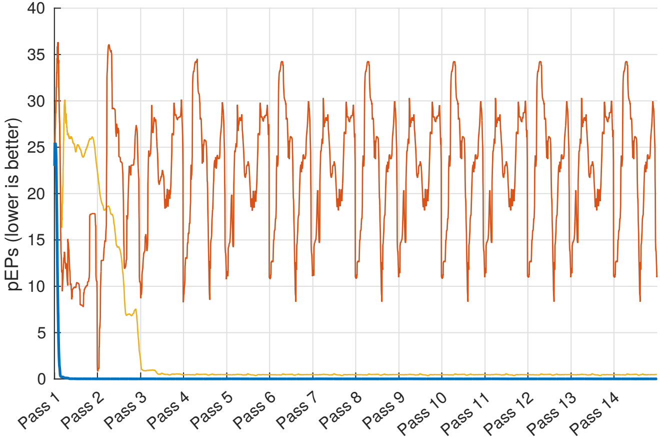
Hypothesis

- The **frame difference** provides the best contribution on average.
- Unique property compared to other algorithms → **temporally memoryless**.

What About the Temporal Stability?

- With **some BGS algorithms**, the background estimation is **never stable**.
- pEPs of frame difference —, VuMeter —, and oracle — (Board sequence).

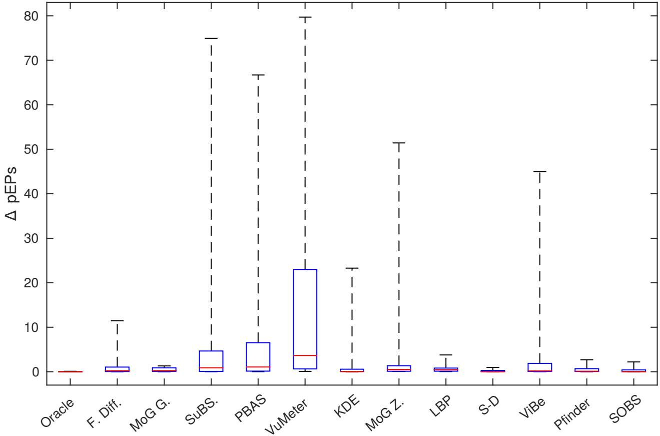
(Laugraud et al. 2017b)



What About the Temporal Stability?

- **Mean variation of the pEPs score** over pass on the SBI sequences.
- SuBSENSE, PBAS, and VuMeter are **not temporally stable**.

(Laugraud et al. 2017b)



Results for SBMnet 2016

Overall | Basic | Intermittent Motion | Clutter | Jitter | Illumination Changes | Background Motion | Very Long | Very Short

Results, all categories combined.

Click on method name for more details.

Method	Average ranking	Average ranking across categories	Average AGE	Average pEPs	Average pCEPS	Average MSSSIM	Average PSNR	CQM
MSCL [15]	1.17	4.88	5.9547	0.0524	0.0171	0.9410	30.8952	31.7049
LaBGen-OF [25]	2.17	4.50	6.1897	0.0566	0.0232	0.9412	29.8957	30.7006
BEWiS [24]	4.33	5.88	6.7094	0.0592	0.0266	0.9282	28.7728	29.6342
LaBGen [6]	4.83	7.75	6.7090	0.0631	0.0265	0.9266	28.6396	29.4668
NExBI [26]	6.00	11.38	6.7778	0.0671	0.0227	0.9196	27.9944	28.8810
LaBGen-P [7]	6.33	8.50	7.0738	0.0706	0.0319	0.9278	28.4660	29.3196
Photomontage [3]	7.17	11.13	7.1950	0.0686	0.0257	0.9189	28.0113	28.8719
SC-SOBS-C4 [9]	8.33	9.63	7.5183	0.0711	0.0242	0.9160	27.6533	28.5601
MAGRPCA [10]	9.50	9.88	8.3132	0.0994	0.0567	0.9401	28.4556	29.3152
Temporal median filter [2]	11.67	9.13	8.2761	0.0984	0.0546	0.9130	27.5364	28.4434
BE-AAPSA [14]	11.67	12.13	7.9086	0.0873	0.0447	0.9127	27.0714	27.9811
Bidirectional Analysis [13]	12.00	11.13	8.3449	0.0756	0.0181	0.9085	26.1722	27.1637
Bidirectional Analysis and Consensus Voting [12]	13.50	13.00	8.5816	0.0724	0.0257	0.9078	26.1018	27.1000

Submitted with $\mathcal{A}$ = Frame difference.

Video Demonstration

URL: https://youtu.be/rYhX8ZizSL0

LaBGen-OF

Motivation

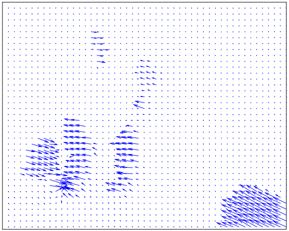
Hypothesis

- The **frame difference** provides the best contribution on average.
- Unique property compared to other algorithms $\rightarrow$ **temporally memoryless**.

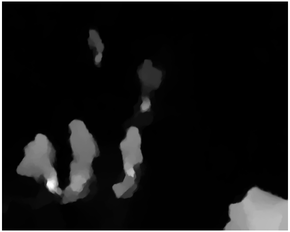
- We checked this hypothesis **experimentally**.
- We shown in a simple case the **impact of temporal memory**.
- We made a **comparison** of motion detection **with or without** temporal memory.
- Motion detectors **with memory** $\rightarrow$ background subtraction.
- Motion detectors **without memory** $\rightarrow$ optical flow, and frame difference.
- To leverage **optical flow**, we made a variant of LaBGen called **LaBGen-OF**.

LaBGen-OF: Motion Detection Step

PETS 2006 - Frame 100



Some velocity vectors (optical flow)



Spatial normalization of ℓ^2-norms

Segmentation map (threshold)

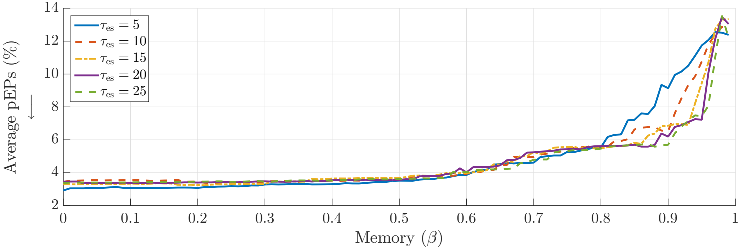
Experience 1: Impact of the Motion Detection Memory on the Performance

- **Average performance** using an **exponential smoothing** BGS on SBI.
- **Background model (**B**)** maintained over **time (**t**)** as follows:

$$B_t = (1-\beta)\cdot F_t + \beta \cdot B_{t-1},$$

  with F being an input frame, and $\beta \in [0,1]$ a parameter.
- **Classification** performed by applying a hard threshold τ_{es} on $|F_t - B_t|$.
- β **increases** $\rightarrow$ **More importance** is given to the **temporal history**.
- First indication that using **no memory** is an **appropriate** choice for LaBGen.

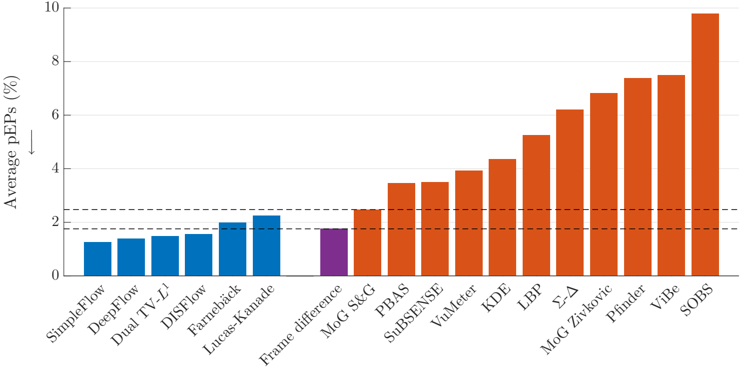
(Laugraud et al. 2017a)



Experience 2: Comparison of Motion Detectors (MD) With or Without Memory

- **Average performance** using several MDs on the SBI dataset.
- The one of MDs **without memory** ■ vary around the one of the **frame diff.** ■.
- All MDs **without memory** ■ are **better** than any MD **with memory** ■.

(Laugraud et al. 2017a)



Results for SBMnet 2016

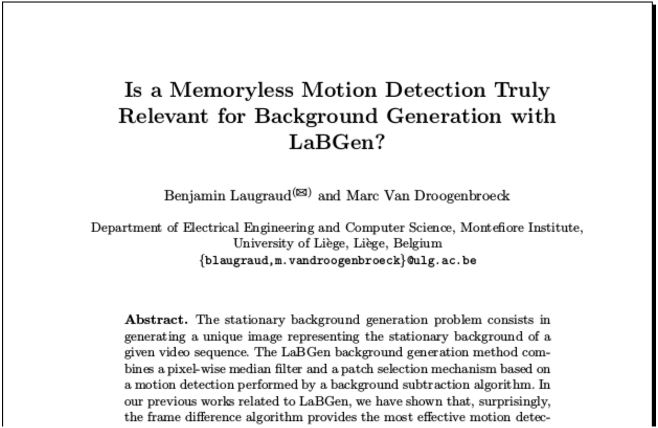
Overall | Basic | Intermittent Motion | Clutter | Jitter | Illumination Changes | Background Motion | Very Long | Very Short

Results, all categories combined.

Click on method name for more details.

Method	Average ranking	Average ranking across categories	Average AGE	Average pEPs	Average pCEPS	Average MSSSIM	Average PSNR	CQM
MSCL [15]	1.17	4.88	5.9547	0.0524	0.0171	0.9410	30.8952	31.7049
LaBGen-OF [25]	2.17	4.50	6.1897	0.0566	0.0232	0.9412	29.8957	30.7006
BEWiS [24]	4.33	5.88	6.7094	0.0592	0.0266	0.9282	28.7728	29.6342
LaBGen [6]	4.83	7.75	6.7090	0.0631	0.0265	0.9266	28.6396	29.4668
NExBI [26]	6.00	11.38	6.7778	0.0671	0.0227	0.9196	27.9944	28.8810
LaBGen-P [7]	6.33	8.50	7.0738	0.0706	0.0319	0.9278	28.4660	29.3196
Photomontage [3]	7.17	11.13	7.1950	0.0686	0.0257	0.9189	28.0113	28.8719
SC-SOBS-C4 [9]	8.33	9.63	7.5183	0.0711	0.0242	0.9160	27.6533	28.5601
MAGRPCA [10]	9.50	9.88	8.3132	0.0994	0.0567	0.9401	28.4556	29.3152
Temporal median filter [2]	11.67	9.13	8.2761	0.0984	0.0546	0.9130	27.5364	28.4434
BE-AAPSA [14]	11.67	12.13	7.9086	0.0873	0.0447	0.9127	27.0714	27.9811
Bidirectional Analysis [13]	12.00	11.13	8.3449	0.0756	0.0181	0.9085	26.1722	27.1637
Bidirectional Analysis and Consensus Voting [12]	13.50	13.00	8.5816	0.0724	0.0257	0.9078	26.1018	27.1000

Submitted with $\mathcal{A} =$ DeepFlow (Weinzaepfel et al. 2013).



Is a Memoryless Motion Detection Truly Relevant for Background Generation with LaBGen?

Benjamin Laugraud(✉) and Marc Van Droogenbroeck

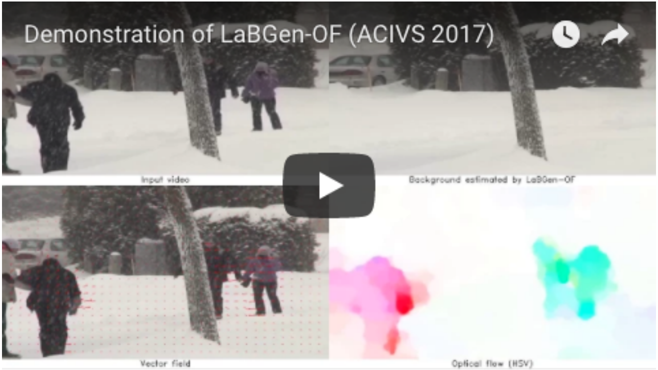
Department of Electrical Engineering and Computer Science, Montefiore Institute, University of Liège, Liège, Belgium
{blaugraud,m.vandroogenbroeck}@ulg.ac.be

Abstract. The stationary background generation problem consists in generating a unique image representing the stationary background of a given video sequence. The LaBGen background generation method combines a pixel-wise median filter and a patch selection mechanism based on a motion detection performed by a background subtraction algorithm. In our previous works related to LaBGen, we have shown that, surprisingly, the frame difference algorithm provides the most effective motion detec-

(Laugraud et al. 2017a)

URL: http://hdl.handle.net/2268/213147

Video Demonstration



URL: https://youtu.be/6tzzY65sCzc

LaBGen-P

- Sometimes, with LaBGen, we have a **“patch effect”**.
- We wanted to make a **pixel-based method** to avoid this effect.
- LaBGen-**P(ixel)**.
- **Example** of background estimated with the **same parameters**:

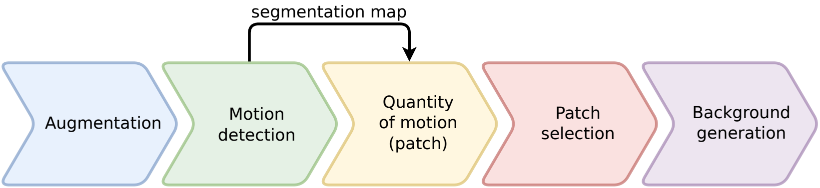
(Laugraud et al. 2016)

LaBGen

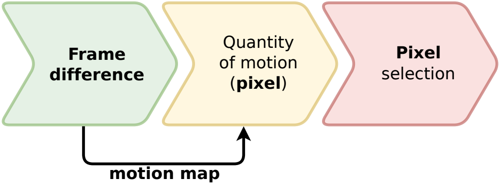
LaBGen-P

Ground-truth

What is New?



⇓ **Modifications performed in LaBGen-P**



Using Motion Maps Instead of Segmentation Maps

- We define ***motion scores*** as raw absolute diff. of intensities (no threshold).
- A ***motion map*** gathers all the motion scores computed between two frames.
- Such a map allows to **capture some shades** about motion.
- For instance: $200 > 20 \rightarrow fg$, $30 > 20 \rightarrow fg$, but $p(fg|200) > p(fg|30)$.
- Avoid to find an **appropriate threshold**!

Motion map

Segmentation map ($\tau = 20$)

Local Estimation of the Quantity of Motion

- Unlike in LaBGen, quantities of motion are **estimated per pixel**.
- However, this estimate take into account the **spatial information**!
- The motion scores available in the **local neighborhood** are **aggregated** (sum).
- The local neighborhood is delimited by a **window centered** on the current pixel.
- The **size of the window** depends on the **parameter** $\mathcal{N}$.

1	2	5	7	5	8	3	3
5	1	8	8	5	2	5	2
6	5	8	3	1	3	6	2
9	3	5	1	1	1	6	1
7	4	2	3	2	1	4	3
7	9	1	1	2	2	1	9
4	8	1	7	7	2	9	8
3	4	2	9	5	6	9	8

Motion map (3×3 window)

quantity of motion of $\blacksquare$ = $\sum_{\blacksquare}$ = 14

Scene Background Modeling Contest (IEEE SBMC 2016)

Results (September 12, 2016)

Overall | Basic | Intermittent Motion | Clutter | Jitter | Illumination Changes | Background Motion | Very Long | Very Short

Results, all categories combined.

Click on method name for more details.

Method	Average ranking	Average ranking across categories	Average AGE	Average pEPs	Average pCEPs	Average MS-SSIM	Average PSNR	Average CQM
LaBGen [6]	2.00	4.75	6.7090	0.0631	0.0265	0.9266	28.6396	29.4668
LaBGen-P [7]	2.83	5.38	7.0738	0.0706	0.0319	0.9278	28.4660	29.3196
Photomontage [3]	3.33	6.50	7.1950	0.0686	0.0257	0.9189	28.0113	28.8719
SC-SOBS-C4 [9]	4.33	6.00	7.5183	0.0711	0.0242	0.9160	27.6533	28.5601
MAGRPCA [10]	5.83	6.88	8.3132	0.0994	0.0567	0.9401	28.4556	29.3152
Temporal median filter [2]	7.17	5.50	8.2761	0.0984	0.0546	0.9130	27.5364	28.4434
BE-AAPSA [14]	7.17	7.88	7.9086	0.0873	0.0447	0.9127	27.0714	27.9811
Bidirectional Analysis [13]	7.67	6.63	8.3449	0.0756	0.0181	0.9085	26.1722	27.1637
Bidirectional Analysis and Consensus Voting [12]	8.67	7.75	8.5816	0.0724	0.0257	0.9078	26.1018	27.1000
TMFG [11]	10.00	6.25	7.4020	0.1051	0.0566	0.9043	27.1347	28.0530
FC-FlowNet [5]	10.17	9.00	9.1131	0.1128	0.0599	0.9162	26.9559	27.8767
RSL2011 [4]	11.17	10.25	9.0443	0.1008	0.0497	0.8891	25.8051	26.7986
AAPSA [1]	12.17	10.88	9.2044	0.1057	0.0523	0.9000	25.3947	26.3021
RMR [8]	12.50	10.00	9.5363	0.1176	0.0582	0.8790	26.5217	27.4549

1. Video for which we would like to define a background image

video/Candela_m1.10.m4v

2. Question

Which background image do you prefer ? --- Please select the correct answer! ---

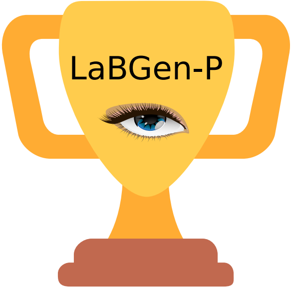
--- Please select the correct answer! ---
I don't know.
The one on the left hand side.
The one on the right hand side.

save the answers and display the next question.

Copyright Piérard Sébastien, 2012

Subjective Evaluation: Web Platform

- **35 human experts** participated to the study.
- **Unable to choose** between LaBGen and LaBGen-P for **38 sequences**.
- **LaBGen-P** was prefered for **26 sequences** and **LaBGen** for **15 sequences**.



The **metrics are not correlated** with the **human eye** and our perception of the background → confirmed recently by Shrotree et al. 2018.


Winner of the Scene Background Modeling Contest
awarded to
B. Laugraud, S. Piérard, and M. Van Droogenbroeck
December 2016
Lucia Maddalena
Pierre-Marc Jodoin

Performance on SBMnet: Average Ranking (07/03/2018)

Results for SBMnet 2016

Overall | Basic | Intermittent Motion | Clutter | Jitter | Illumination Changes | Background Motion | Very Long | Very Short

Results, all categories combined.

Click on method name for more details.

Method	Average ranking	Average ranking across categories	Average AGE	Average pEPs	Average pCEPS	Average MSSSIM	Average PSNR	CQM
MSCL [15]	1.17	4.88	5.9547	0.0524	0.0171	0.9410	30.8952	31.7049
LaBGen-OF [25]	2.17	4.50	6.1897	0.0566	0.0232	0.9412	29.8957	30.7006
BEWiS [24]	4.33	5.88	6.7094	0.0592	0.0266	0.9282	28.7728	29.6342
LaBGen [6]	4.83	7.75	6.7090	0.0631	0.0265	0.9266	28.6396	29.4668
NExBI [26]	6.00	11.38	6.7778	0.0671	0.0227	0.9196	27.9944	28.8810
LaBGen-P [7]	6.33	8.50	7.0738	0.0706	0.0319	0.9278	28.4660	29.3196
Photomontage [3]	7.17	11.13	7.1950	0.0686	0.0257	0.9189	28.0113	28.8719
SC-SOBS-C4 [9]	8.33	9.63	7.5183	0.0711	0.0242	0.9160	27.6533	28.5601
MAGRPCA [10]	9.50	9.88	8.3132	0.0994	0.0567	0.9401	28.4556	29.3152
Temporal median filter [2]	11.67	9.13	8.2761	0.0984	0.0546	0.9130	27.5364	28.4434
BE-AAPSA [14]	11.67	12.13	7.9086	0.0873	0.0447	0.9127	27.0714	27.9811
Bidirectional Analysis [13]	12.00	11.13	8.3449	0.0756	0.0181	0.9085	26.1722	27.1637
Bidirectional Analysis and Consensus Voting [12]	13.50	13.00	8.5816	0.0724	0.0257	0.9078	26.1018	27.1000

LaBGen-P: A Pixel-Level Stationary Background Generation Method Based on LaBGen

Benjamin Laugraud, Sébastien Piérard, and Marc Van Droogenbroeck
INTELSIG Laboratory, University of Liège, Belgium
{BLaugraud, Sebastien.Pierard, M.VanDroogenbroeck}@ulg.ac.be

***Abstract*—Estimating the stationary background of a video sequence is useful in many applications like surveillance, segmentation, compression, inpainting, privacy protection, and computational photography. To perform this task, we introduce the LaBGen-P method based on the principles of LaBGen and the conclusions drawn in the corresponding paper. It combines a pixel-wise median filter and a pixel selection mechanism based on a motion detection performed by the frame difference algorithm. By working with pixels instead of patches, as originally done in LaBGen, it avoids some discontinuities between different spatial areas and generates better visual results. In this paper, we describe the LaBGen-P method, study its performance on the**

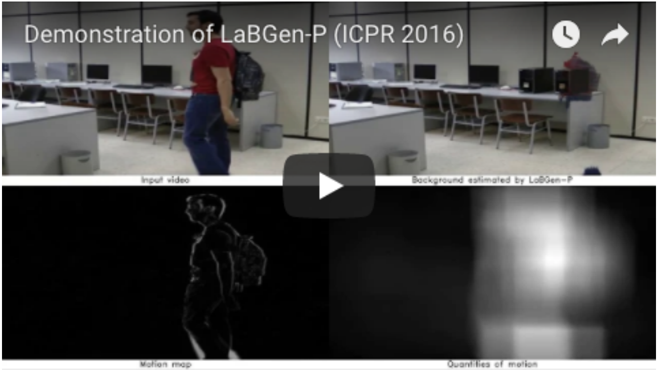
LaBGen | LaBGen-P | Ground truth

Fig. 1. Background estimated for the Board sequence of the SBMnet dataset by LaBGen and LaBGen-P with the parameters discussed in Section III-C. The discontinuities observable between different spatial areas make the estimation visually incoherent. They are avoided with the LaBGen-P method.

(Laugraud et al. 2016)

URL: http://hdl.handle.net/2268/201146

Video Demonstration



URL: https://youtu.be/lcXHM42EeZo

Unsolved Issues

Intermittent motion

Strong illumination changes

- Intermittent motion → **memoryless** motion detection **inefficient**.
- Strong illum. changes → **no intrinsic mechanism** for temporal consistency.

Conclusion

Conclusion

- A **lot of challenges** make background initialization difficult to solve.
- Methods based on various math. tools $\rightarrow$ **none of them** solves all **challenges**.
- **LaBGen** is a state-of-the art method based on **temporal median filtering**.
- Observations considered as in **motion are evicted** from the median “buffer”.
- Results of motion detection **combined spatially**.
- Motion detectors **without temporal memory** perform the best for LaBGen.
- According to state-of-the-art datasets, LaBGen is **among the top performers**.
- **Robustness** against illum. changes and inter. motion still **to be improved**.
- The **human eye** sometimes disagree with the current **evalution methodology**.

Perspectives

- Create new and **innovative methods**.

  Welcome in our field :-)!

- Improve the robustness of **LaBGen** against **illumination changes**.

  We have nothing in our pipeline for the moment!

- Improve the robustness of **LaBGen** against **intermittent motion**.

  We are currently working on something!

- Improve the **evaluation methodology**.

  Recently, a new metric has been proposed for background initialization (Shrotree et al. 2018).

- Deep learning needs **a lot of data**.

  A massively huge dataset full of ground-truth would be really appreciated :-)!

شكراً جزيلاً
thank you
danke
謝謝
ngiyabonga
teşekkür ederim
gracias
merci
obrigado
go raibh maith agat
mochchakkeram
sukriya
kop khun krap
terima kasih
arigatô
dakujem
grazie
takk
dziękuję
bedankt
hvala
spasibo
спасибо
Баярлалаа
рахмат
dank je
tapadh leat
감사합니다
ありがとう
ευχαριστώ
мерси
хвала
дякую
спасибі
sagolun
taiku
chnorakaloutioun
gratias ago
gràcies
blagodaram
misaotra
matondo
paldies
grazzi
mahalo
asante
manana
obrigada
tenki
murakoze
chokrane
mamnun
djiere dieuf
tau
sulpáy
trugarez
shukriya
mercê
dhanyavadagalu
diolch
rahmet
tanemirt
xièxie
najis tuke
kam sah hamnida
rahmat
didi madloba
mèsi
dêkuji
sobodi
enkosi
bayarlalaa
gràcie
nanni
nandri
kiitos
dankie
dhanyavad
maururu
köszönöm
faafetai lava
vinaka
mersi
kia ora
barka
welalin
tack
spas
aciu
akun dankon
তোমাকে ধন্যবাদ

References

References

[Aga+04] A. Agarwala et al. "Interactive Digital Photomontage." In: *ACM Transactions on Graphics* 23.3 (2004), pp. 294–302. DOI: 10.1145/1015706.1015718.

[Bar+11] O. Barnich and M. Van Droogenbroeck. "ViBe: A universal background subtraction algorithm for video sequences." In: *IEEE Transactions on Image Processing* 20.6 (2011), pp. 1709–1724. DOI: 10.1109/TIP.2010.2101613.

[Bis06] C. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, 2006.

[Bou+17] T. Bouwmans, L. Maddalena, and A. Petrosino. "Scene Background Initialization: a Taxonomy." In: *Pattern Recognition Letters* 96 (2017), pp. 3–11. DOI: 10.1016/j.patrec.2016.12.024.

[Jod+17] P.-M. Jodoin et al. "Extensive Benchmark and Survey of Modeling Methods for Scene Background Initialization." In: *IEEE Transactions on Image Processing* 26.11 (2017), pp. 5244–5256. DOI: 10.1109/TIP.2017.2728181.

References

[Lau+15] B. Laugraud et al. “Simple median-based method for stationary background generation using background subtraction algorithms.” In: *International Conference on Image Analysis and Processing (ICIAP), Workshop on Scene Background Modeling and Initialization (SBMI)*. Vol. 9281. Lecture Notes in Computer Science. Genova, Italy: Springer, 2015, pp. 477–484. DOI: 10.1007/978-3-319-23222-5_58.

[Lau+16] B. Laugraud, S. Piérard, and M. Van Droogenbroeck. “LaBGen-P: A Pixel-Level Stationary Background Generation Method Based on LaBGen.” In: *IEEE International Conference on Pattern Recognition (ICPR), IEEE Scene Background Modeling Contest (SBMC)*. Cancún, Mexico, 2016, pp. 107–113. DOI: 10.1109/ICPR.2016.7899617.

[Lau+17a] B. Laugraud and M. Van Droogenbroeck. “Is a Memoryless Motion Detection Truly Relevant for Background Generation with LaBGen?” In: *Advanced Concepts for Intelligent Vision Systems (ACIVS)*. Vol. 10617. Lecture Notes in Computer Science. Antwerp, Belgium: Springer, 2017, pp. 443–454. DOI: 10.1007/978-3-319-70353-4_38.

References

[Lau+17b] B. Laugraud, S. Piérard, and M. Van Droogenbroeck. "LaBGen: A method based on motion detection for generating the background of a scene." In: *Pattern Recognition Letters* 96 (2017), pp. 12–21. DOI: 10.1016/j.patrec.2016.11.022.

[Mad+15] L. Maddalena and A. Petrosino. "Towards Benchmarking Scene Background Initialization." In: *International Conference on Image Analysis and Processing Workshops (ICIAP Workshops)*. Vol. 9281. Lecture Notes in Computer Science. 2015, pp. 469–476. DOI: 10.1007/978-3-319-23222-5_57.

[Min+16] T. Minematsu, A. Shimada, and R.-I. Taniguchi. "Background Initialization based on Bidirectional Analysis and Consensus Voting." In: *IEEE International Conference on Pattern Recognition (ICPR), IEEE Scene Background Modeling Contest (SBMC)*. Cancún, Mexico, 2016, pp. 126–131. DOI: 10.1109/ICPR.2016.7899620.

[Ort+16] D. Ortego, J. M. SanMiguel, and J. M. Martínez. "Rejection based multipath reconstruction for background estimation in video sequences with stationary objects." In: *Computer Vision and Image Understanding* 147 (2016), pp. 23–37. DOI: 10.1016/j.cviu.2016.03.012.

[RA+17] G. Ramirez-Alonso, J. A. Ramirez-Quintana, and M. I. Chacon-Murguia. “Temporal weighted learning model for background estimation with an automatic re-initialization stage and adaptive parameters update.” In: *Pattern Recognition Letters* 96 (2017), pp. 34–44. DOI: 10.1016/j.patrec.2017.01.011.

[Red+11] V. Reddy, C. Sanderson, and B. Lovell. “A Low-Complexity Algorithm for Static Background Estimation from Cluttered Image Sequences in Surveillance Contexts.” In: *EURASIP Journal on Image and Video Processing* 2011 (2011), p. 164956. DOI: 10.1155/2011/164956.

[Shr+18] A. Shrotree and L. Karam. “Full Reference Objective Quality Assessment for Recontructed Background Images.” In: *CoRR* abs/1803.04103 (2018).

[Sob+17] A. Sobral and E.-H. Zahzah. “Matrix and tensor completion algorithms for background model initialization: A comparative evaluation.” In: *Pattern Recognition Letters* 96 (2017), pp. 22–33. DOI: 10.1016/j.patrec.2016.12.019.

[Sta+99] C. Stauffer and E. Grimson. “Adaptive background mixture models for real-time tracking.” In: *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 2. Fort Collins, Colorado, USA, 1999, pp. 246–252. DOI: 10.1109/CVPR.1999.784637.

[Wan+03] Z. Wang, E. P. Simoncelli, and A. C. Bovik. "Multiscale structural similarity for image quality assessment." In: *Asilomar Conference on Signals, Systems and Computers*. Vol. 2. Pacific Grove, California, USA, 2003, pp. 1398–1402. DOI: 10.1109/ACSSC.2003.1292216.

[Wan+04] Z. Wang et al. "Image quality assessment: from error visibility to structural similarity." In: *IEEE Transactions on Image Processing* 13.4 (2004), pp. 600–612. DOI: 10.1109/TIP.2003.819861.

[Wan+06] H. Wang and D. Suter. "A Novel Robust Statistical Method for Background Initialization and Visual Surveillance." In: *Asian Conference on Computer Vision (ACCV)*. Vol. 3851. Lecture Notes in Computer Science. Berlin, Heidelberg, 2006, pp. 328–337. DOI: 10.1007/11612032_34.

[Wei+13] P. Weinzaepfel et al. "DeepFlow: Large displacement optical flow with deep matching." In: *International Conference on Computer Vision (ICCV)*. Sydney, Australia, 2013, pp. 1385–1392. DOI: 10.1109/ICCV.2013.175.

References

[Yal+13] Y. Yalman and I. Ertürk. "A new color image quality measure based on YUV transformation and PSNR for human vision system." In: *Turkish Journal of Electrical Engineering & Computer Sciences* 21.2 (2013), pp. 603–613. DOI: 10.3906/ELK-1111-11.

[De +17] M. De Gregorio and M. Giordano. "Background estimation by weightless neural networks." In: *Pattern Recognition Letters* 96 (2017), pp. 55–65. DOI: 10.1016/j.patrec.2017.05.029.