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**

![Example Images]
What is Background Initialization?

- Several applications → video surveillance, computational photography, etc (see Maddalena et al. 2015).

- Not as easy as it looks → 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](image1)
![Input frame](image2)
![Expected output](image3)
Intermittent Motion Challenge

- Objects that **stop for a short while**.

- Objects that are **abandoned**.

- Background objects **starting to move**.

Temporal median  
LaBGen  
(Reugraud 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**.
Very Short Challenge

- 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)
Methods Based on Temporal Statistics

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.).
Methods Based on Temporal Statistics

**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 \( R_t^s \) at time \( t \).
- In each spatial area \( s \) \( \rightarrow \text{motion filtering} \) to discard patches in motion (f. diff.).
- **Custering** performed to build a set of candidates \( C_i^s \).
- Several numbers of clusters \( N^s \) tested \( \rightarrow \text{between 1 and the number of SSIs} \) detected during motion filtering.
- Choose \( N^s \) \( \rightarrow \text{metric max. compactess and separation, and min. similarity} \).
- Candidates \( C_i^s \) in spatial area \( s \) \( \rightarrow \text{mean} \) of each cluster \( K_i^s \), with \( l = 1, \ldots, N^s \).
- **Seed selection** selects **highly reliable candidates** \( S^s \) with a large cluster cardinality and a low motion activity in the associated spatial area.
- Spatial areas \( s \) with **empty seeds** \( S^s = \emptyset \) 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$^p$ weightless neural network.
- Learning from the sequence without any annotation → unsupervised.
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 → reward $\rho$.

Other cells → decreased by a punishment $\Psi$.

Background recovered from the addresses of the cells with the highest values.
There are two major reviews!

(Bouwmans et al. 2017)
URL: https://doi.org/10.1016/j.patrec.2016.12.024

(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](image1.png) ![Ground-truth](image2.png)

![Abs. Difference](image3.png) ![EPs](image4.png) ![CEPs](image5.png)
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):

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

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.
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 → 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**.

### 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](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.
### Results for SBMnet 2016

**Click on method name for more details.**

<table>
<thead>
<tr>
<th>Method</th>
<th>Average ranking</th>
<th>Average ranking across categories</th>
<th>Average AGE</th>
<th>Average pEPs</th>
<th>Average pCEPS</th>
<th>Average MSSSIM</th>
<th>Average PSNR</th>
<th>CQM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMSL [15]</td>
<td>1.17</td>
<td>4.88</td>
<td>5.9547</td>
<td>0.0524</td>
<td>0.0171</td>
<td>0.9410</td>
<td>30.8952</td>
<td>31.7040</td>
</tr>
<tr>
<td>LaBGen QE [20]</td>
<td>2.17</td>
<td>4.50</td>
<td>6.1897</td>
<td>0.0566</td>
<td>0.0232</td>
<td>0.9412</td>
<td>29.8957</td>
<td>30.7006</td>
</tr>
<tr>
<td>BEWS [24]</td>
<td>4.33</td>
<td>5.88</td>
<td>6.7004</td>
<td>0.0592</td>
<td>0.0266</td>
<td>0.9282</td>
<td>28.7728</td>
<td>29.6342</td>
</tr>
<tr>
<td>LaBGen [6]</td>
<td>4.83</td>
<td>7.75</td>
<td>6.7006</td>
<td>0.0631</td>
<td>0.0265</td>
<td>0.9256</td>
<td>28.6306</td>
<td>29.4683</td>
</tr>
<tr>
<td>NGSBi [28]</td>
<td>5.00</td>
<td>11.38</td>
<td>6.7778</td>
<td>0.0671</td>
<td>0.0227</td>
<td>0.9106</td>
<td>27.9944</td>
<td>28.8810</td>
</tr>
<tr>
<td>LaBGen-P [7]</td>
<td>6.33</td>
<td>8.50</td>
<td>7.0738</td>
<td>0.0706</td>
<td>0.0319</td>
<td>0.9278</td>
<td>28.4660</td>
<td>29.3196</td>
</tr>
<tr>
<td>Photomontage [3]</td>
<td>7.17</td>
<td>11.13</td>
<td>7.1950</td>
<td>0.0686</td>
<td>0.0257</td>
<td>0.9189</td>
<td>28.0113</td>
<td>28.8719</td>
</tr>
<tr>
<td>SC-SOBS-C4 [9]</td>
<td>8.33</td>
<td>9.63</td>
<td>7.5183</td>
<td>0.0711</td>
<td>0.0242</td>
<td>0.9160</td>
<td>27.6533</td>
<td>28.5601</td>
</tr>
<tr>
<td>MAGRPCA [10]</td>
<td>9.50</td>
<td>9.88</td>
<td>8.3132</td>
<td>0.0994</td>
<td>0.0567</td>
<td>0.9401</td>
<td>28.4585</td>
<td>29.3152</td>
</tr>
<tr>
<td>Temporal median filter [2]</td>
<td>11.67</td>
<td>9.13</td>
<td>8.2761</td>
<td>0.0984</td>
<td>0.0546</td>
<td>0.9130</td>
<td>27.5364</td>
<td>28.4434</td>
</tr>
</tbody>
</table>
SBMnet: Category Average Ranking (Clutter)

Results for SBMnet 2016

Click on method name for more details.

<table>
<thead>
<tr>
<th>Method</th>
<th>Average ranking</th>
<th>Average AGE</th>
<th>Average pLEPs</th>
<th>Average pCELPS</th>
<th>Average MSSSIM</th>
<th>Average PSNR</th>
<th>CQM</th>
</tr>
</thead>
<tbody>
<tr>
<td>LaBGen-OE [25]</td>
<td>1.17</td>
<td>4.1821</td>
<td>0.0246</td>
<td>0.0117</td>
<td>0.9640</td>
<td>32.6339</td>
<td>33.4664</td>
</tr>
<tr>
<td>MSCL [15]</td>
<td>1.83</td>
<td>5.2695</td>
<td>0.0275</td>
<td>0.0094</td>
<td>0.9629</td>
<td>31.3743</td>
<td>32.2837</td>
</tr>
<tr>
<td>NExBi [29]</td>
<td>3.33</td>
<td>5.3001</td>
<td>0.0414</td>
<td>0.0141</td>
<td>0.9379</td>
<td>30.0056</td>
<td>30.0847</td>
</tr>
<tr>
<td>Photomontage [3]</td>
<td>6.17</td>
<td>6.8195</td>
<td>0.0543</td>
<td>0.0294</td>
<td>0.8802</td>
<td>28.5554</td>
<td>29.4882</td>
</tr>
<tr>
<td>Bidirectional Analysis [33]</td>
<td>8.57</td>
<td>6.6565</td>
<td>0.0497</td>
<td>0.0177</td>
<td>0.9243</td>
<td>26.4376</td>
<td>27.5267</td>
</tr>
<tr>
<td>SC-SOBS-C4 [9]</td>
<td>7.33</td>
<td>7.0500</td>
<td>0.0644</td>
<td>0.0304</td>
<td>0.8939</td>
<td>28.0077</td>
<td>29.0737</td>
</tr>
<tr>
<td>MAGRPCA [10]</td>
<td>7.57</td>
<td>8.1589</td>
<td>0.0647</td>
<td>0.0294</td>
<td>0.9446</td>
<td>26.6872</td>
<td>27.5988</td>
</tr>
<tr>
<td>RWavelets [20]</td>
<td>7.67</td>
<td>8.1589</td>
<td>0.0647</td>
<td>0.0294</td>
<td>0.9446</td>
<td>26.6872</td>
<td>27.5988</td>
</tr>
<tr>
<td>RSL2011 [4]</td>
<td>8.33</td>
<td>7.3013</td>
<td>0.0701</td>
<td>0.0375</td>
<td>0.9067</td>
<td>27.9304</td>
<td>28.0763</td>
</tr>
<tr>
<td>LaBGen-P [7]</td>
<td>8.33</td>
<td>7.8947</td>
<td>0.0886</td>
<td>0.0678</td>
<td>0.8967</td>
<td>28.1140</td>
<td>29.1305</td>
</tr>
</tbody>
</table>
### SBMnet: Method Details (LaBGen)

**Contact name**: Benjamin Laugraud  
**Contact email**: blaugraud@ulg.ac.be  
**Contact university/company**: University of Liège  
**Method’s name**: LaBGen  
**Reference**: B. Laugraud, S. Plérand, 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](http://www.telecom.ulg.ac.be/labgen)  

**Parameters**: $A = \text{frame}_\text{difference}, P = 1, N = 3, S = 19$

You can download the background estimation results of this method by clicking [here](http://www.telecom.ulg.ac.be/labgen).

<table>
<thead>
<tr>
<th>Video</th>
<th>AGE</th>
<th>pEPs</th>
<th>pCEPS</th>
<th>MSSSIM</th>
<th>PSNR</th>
<th>CQM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blurred</td>
<td>1.0990</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.9975</td>
<td>41.5779</td>
<td>41.5641</td>
</tr>
<tr>
<td>PETS2006</td>
<td>2.0432</td>
<td>0.0027</td>
<td>0.0019</td>
<td>0.9896</td>
<td>35.2304</td>
<td>35.9487</td>
</tr>
<tr>
<td>Intersection</td>
<td>2.4285</td>
<td>0.0009</td>
<td>0.0000</td>
<td>0.9900</td>
<td>37.2576</td>
<td>37.7342</td>
</tr>
</tbody>
</table>
LaBGen
Ideas

1. Median buffer **polluted with foreground** elements → 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

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.

---

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

Benjamin Laugraud, Sébastien Piérard, Marc Braham, and Marc Van Droogenbroeck

INTEGRLE 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 removes the set of color candidates, for each pixel, before the median is applied. Our method is simple and fully generic as

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

(Laugraud et al. 2015)
Step 1: Augmentation

- BGS needs sometimes a **long training** → problem with **short sequences**.

- The augmentation step **increases the duration** of the input video sequence.

- In practice, we process the sequence in $P$ **passes**.

- An **odd** (resp. **even**) pass is performed **forwards** (resp. **backwards**).
ViBe (Barnich et al. 2011) is highly sensitive to bootstrap.

First frame contains several foreground objects → ghosts!

The augmentation mechanism can help to reduce false alarms.

(Laugraud et al. 2017b)
Step 2: Motion Detection

- The BGS algorithm being used is the parameter $A$.

- LaBGen does not leverage the model of $A$, 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 $N \times N$ 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, \( S \) patches are selected.

- The \( S \) 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 $S$ selected patches.

- The **background** is then **generated**.
Summary of the Pipeline

(Laugaoud et al. 2017a)

1. Augmentation step
2. Motion detection step
3. Estimation step
4. Selection step
5. Generation step (output)
What is the Best BGS Algorithm for our Framework?

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

Click on method name for more details.

<table>
<thead>
<tr>
<th>Method</th>
<th>Average ranking</th>
<th>Average ranking across categories</th>
<th>Average AGE</th>
<th>Average pEPs</th>
<th>Average pCEPS</th>
<th>Average MSSIM</th>
<th>Average PSNR</th>
<th>CDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSCL [15]</td>
<td>1.17</td>
<td>4.68</td>
<td>5.9547</td>
<td>0.0524</td>
<td>0.0171</td>
<td>0.9410</td>
<td>30.8952</td>
<td>31.7049</td>
</tr>
<tr>
<td>LaBGen-OF [25]</td>
<td>2.17</td>
<td>4.50</td>
<td>6.1897</td>
<td>0.0566</td>
<td>0.0232</td>
<td>0.9412</td>
<td>29.8957</td>
<td>30.7006</td>
</tr>
<tr>
<td>BEVIS [24]</td>
<td>4.33</td>
<td>5.88</td>
<td>6.7094</td>
<td>0.0592</td>
<td>0.0286</td>
<td>0.9282</td>
<td>28.7728</td>
<td>29.6342</td>
</tr>
<tr>
<td>LaBGen [6]</td>
<td>4.83</td>
<td>7.75</td>
<td>6.7090</td>
<td>0.0631</td>
<td>0.0285</td>
<td>0.9266</td>
<td>28.6396</td>
<td>29.4688</td>
</tr>
<tr>
<td>NEExp [26]</td>
<td>6.00</td>
<td>11.38</td>
<td>6.7778</td>
<td>0.0671</td>
<td>0.0227</td>
<td>0.9166</td>
<td>27.9944</td>
<td>28.3810</td>
</tr>
<tr>
<td>LaBGen-P [7]</td>
<td>6.33</td>
<td>8.50</td>
<td>7.0738</td>
<td>0.0706</td>
<td>0.0319</td>
<td>0.9278</td>
<td>29.4650</td>
<td>29.3106</td>
</tr>
<tr>
<td>Photometry [3]</td>
<td>7.17</td>
<td>11.13</td>
<td>7.1950</td>
<td>0.0666</td>
<td>0.0257</td>
<td>0.9169</td>
<td>29.0113</td>
<td>28.8719</td>
</tr>
<tr>
<td>SC-SOBS-C4 [6]</td>
<td>6.33</td>
<td>9.63</td>
<td>7.5163</td>
<td>0.0711</td>
<td>0.0242</td>
<td>0.9160</td>
<td>27.0533</td>
<td>26.5601</td>
</tr>
<tr>
<td>MAGRPGA [10]</td>
<td>5.50</td>
<td>9.68</td>
<td>8.3132</td>
<td>0.0954</td>
<td>0.0507</td>
<td>0.9401</td>
<td>20.4556</td>
<td>29.3152</td>
</tr>
<tr>
<td>Temporal median filter [2]</td>
<td>11.67</td>
<td>9.13</td>
<td>8.2761</td>
<td>0.0584</td>
<td>0.0546</td>
<td>0.9130</td>
<td>27.5354</td>
<td>28.4434</td>
</tr>
<tr>
<td>BF-AAPS [14]</td>
<td>11.67</td>
<td>12.13</td>
<td>7.9086</td>
<td>0.0873</td>
<td>0.0447</td>
<td>0.9127</td>
<td>27.0714</td>
<td>27.9811</td>
</tr>
<tr>
<td>Bidirectional Analysis [13]</td>
<td>12.00</td>
<td>11.13</td>
<td>8.3449</td>
<td>0.0756</td>
<td>0.0181</td>
<td>0.9085</td>
<td>26.1722</td>
<td>27.1837</td>
</tr>
<tr>
<td>Bidirectional Analysis and Consensus Voting [12]</td>
<td>13.80</td>
<td>13.00</td>
<td>8.5816</td>
<td>0.0724</td>
<td>0.0257</td>
<td>0.9078</td>
<td>26.1018</td>
<td>27.1000</td>
</tr>
</tbody>
</table>

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 → **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** → background subtraction.
■ Motion detectors **without memory** → 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 $\ell^2$-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 \(\tau_{es}\) on \(|F_t - B_t|\).

- **\(\beta\) 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)
Performance on SBMnet: Average Ranking (07/03/2018)

## Results for SBMnet 2016

<table>
<thead>
<tr>
<th>Method</th>
<th>Average ranking</th>
<th>Average ranking across categories</th>
<th>Average AGF</th>
<th>Average pEPs</th>
<th>Average pCEPS</th>
<th>Average MSSIM</th>
<th>Average PSNR</th>
<th>CQM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSGL [15]</td>
<td>1.17</td>
<td>4.68</td>
<td>5.9547</td>
<td>0.0524</td>
<td>0.0171</td>
<td>0.9410</td>
<td>30.8952</td>
<td>31.7049</td>
</tr>
<tr>
<td>LaBGen-OF [28]</td>
<td>2.17</td>
<td>4.50</td>
<td>6.1897</td>
<td>0.0565</td>
<td>0.0232</td>
<td>0.9412</td>
<td>29.8857</td>
<td>30.7006</td>
</tr>
<tr>
<td>REVIS [24]</td>
<td>4.33</td>
<td>5.88</td>
<td>6.7094</td>
<td>0.0552</td>
<td>0.0286</td>
<td>0.9262</td>
<td>28.7728</td>
<td>29.6342</td>
</tr>
<tr>
<td>LaBGen [6]</td>
<td>4.83</td>
<td>7.75</td>
<td>6.7390</td>
<td>0.0631</td>
<td>0.0285</td>
<td>0.9266</td>
<td>28.8396</td>
<td>29.4688</td>
</tr>
<tr>
<td>NEIRI [26]</td>
<td>6.00</td>
<td>11.38</td>
<td>6.7778</td>
<td>0.0671</td>
<td>0.0227</td>
<td>0.9166</td>
<td>27.9944</td>
<td>28.8100</td>
</tr>
<tr>
<td>LaBGen-P [7]</td>
<td>6.33</td>
<td>8.50</td>
<td>7.0738</td>
<td>0.0706</td>
<td>0.0319</td>
<td>0.9278</td>
<td>28.4880</td>
<td>29.3166</td>
</tr>
<tr>
<td>Photomontage [3]</td>
<td>7.17</td>
<td>11.13</td>
<td>7.1950</td>
<td>0.0666</td>
<td>0.0257</td>
<td>0.9169</td>
<td>28.0113</td>
<td>28.8719</td>
</tr>
<tr>
<td>SC-SOBS-C4 [8]</td>
<td>6.33</td>
<td>9.63</td>
<td>7.5183</td>
<td>0.0711</td>
<td>0.0242</td>
<td>0.9160</td>
<td>27.0533</td>
<td>28.3601</td>
</tr>
<tr>
<td>MAGRPCA [10]</td>
<td>5.50</td>
<td>9.68</td>
<td>8.3132</td>
<td>0.0594</td>
<td>0.0507</td>
<td>0.9401</td>
<td>29.4556</td>
<td>29.3152</td>
</tr>
<tr>
<td>Temporal median filter [2]</td>
<td>11.67</td>
<td>9.13</td>
<td>8.2761</td>
<td>0.0964</td>
<td>0.0546</td>
<td>0.9130</td>
<td>27.5354</td>
<td>28.4434</td>
</tr>
<tr>
<td>BE-AAPSA [14]</td>
<td>11.67</td>
<td>12.13</td>
<td>7.9086</td>
<td>0.0873</td>
<td>0.0447</td>
<td>0.9127</td>
<td>27.0714</td>
<td>27.9811</td>
</tr>
<tr>
<td>Bidirectional Analysis [13]</td>
<td>12.00</td>
<td>11.13</td>
<td>8.3449</td>
<td>0.0756</td>
<td>0.0181</td>
<td>0.9085</td>
<td>26.1722</td>
<td>27.1637</td>
</tr>
<tr>
<td>Bidirectional Analysis and Consensus Voting [12]</td>
<td>13.50</td>
<td>13.00</td>
<td>8.5816</td>
<td>0.0724</td>
<td>0.0257</td>
<td>0.9078</td>
<td>26.1018</td>
<td>27.1000</td>
</tr>
</tbody>
</table>

Submitted with $\mathcal{A} = \text{DeepFlow}$ (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
URL: https://youtu.be/6tzzY65sCzc
LaBGen-P
Motivation

- 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)
What is New?

Modifications performed in LaBGen-P
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**!
Local Estimation of the Quantity of Motion

- Unlike in LaBGen, quantities of motion are estimated per pixel.
- However, this estimate takes 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 $N$.

```
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 $\blacksquare = \sum = 14$
## Scene Background Modeling Contest (IEEE SBMC 2016)

### Results (September 12, 2016)

Results, all categories combined.

**Click on method name for more details.**

<table>
<thead>
<tr>
<th>Method</th>
<th>Average ranking</th>
<th>Average ranking across categories</th>
<th>Average AGE</th>
<th>Average pEPs</th>
<th>Average pCEPs</th>
<th>Average MS-SSIM</th>
<th>Average PSNR</th>
<th>Average CQM</th>
</tr>
</thead>
<tbody>
<tr>
<td>LaBGen [6]</td>
<td>2.00</td>
<td>4.75</td>
<td>6.7090</td>
<td>0.0631</td>
<td>0.0255</td>
<td>0.9266</td>
<td>28.6396</td>
<td>29.4668</td>
</tr>
<tr>
<td>LaBGen-P [7]</td>
<td>2.83</td>
<td>5.38</td>
<td>7.0738</td>
<td>0.0706</td>
<td>0.0319</td>
<td>0.9278</td>
<td>28.4650</td>
<td>29.3106</td>
</tr>
<tr>
<td>Photonorstige [3]</td>
<td>3.33</td>
<td>6.50</td>
<td>7.1950</td>
<td>0.0666</td>
<td>0.0257</td>
<td>0.9169</td>
<td>28.0113</td>
<td>28.6719</td>
</tr>
<tr>
<td>SC SOBS-C4 [9]</td>
<td>4.33</td>
<td>6.00</td>
<td>7.5183</td>
<td>0.0711</td>
<td>0.0242</td>
<td>0.9160</td>
<td>27.6533</td>
<td>28.5601</td>
</tr>
<tr>
<td>MAGRPCA [10]</td>
<td>5.63</td>
<td>6.83</td>
<td>8.3132</td>
<td>0.0954</td>
<td>0.0567</td>
<td>0.9401</td>
<td>28.4556</td>
<td>29.3152</td>
</tr>
<tr>
<td>Temporal median filter [2]</td>
<td>7.17</td>
<td>5.50</td>
<td>8.2761</td>
<td>0.0984</td>
<td>0.0546</td>
<td>0.9130</td>
<td>27.5364</td>
<td>28.4434</td>
</tr>
<tr>
<td>BE-AAPSA [14]</td>
<td>7.17</td>
<td>7.33</td>
<td>7.9006</td>
<td>0.0873</td>
<td>0.0447</td>
<td>0.9127</td>
<td>27.0714</td>
<td>27.9611</td>
</tr>
<tr>
<td>Bidirectional Analysis [13]</td>
<td>7.67</td>
<td>6.83</td>
<td>8.3419</td>
<td>0.0756</td>
<td>0.0181</td>
<td>0.9086</td>
<td>26.1722</td>
<td>27.1637</td>
</tr>
<tr>
<td>Bidirectional Analysis and Consensus Voting [12]</td>
<td>8.67</td>
<td>7.75</td>
<td>8.5016</td>
<td>0.0724</td>
<td>0.0257</td>
<td>0.9070</td>
<td>26.1013</td>
<td>27.1000</td>
</tr>
<tr>
<td>TMFG [11]</td>
<td>10.00</td>
<td>6.25</td>
<td>7.4020</td>
<td>0.1051</td>
<td>0.0588</td>
<td>0.9043</td>
<td>27.1347</td>
<td>28.0650</td>
</tr>
<tr>
<td>FC-FlowNet [5]</td>
<td>10.17</td>
<td>9.00</td>
<td>9.1131</td>
<td>0.1128</td>
<td>0.0539</td>
<td>0.9162</td>
<td>26.9559</td>
<td>27.8767</td>
</tr>
<tr>
<td>AAPSA [11]</td>
<td>12.17</td>
<td>10.88</td>
<td>9.2044</td>
<td>0.1057</td>
<td>0.0523</td>
<td>0.9000</td>
<td>25.3947</td>
<td>26.3021</td>
</tr>
<tr>
<td>RMR [8]</td>
<td>12.50</td>
<td>10.00</td>
<td>9.5363</td>
<td>0.1176</td>
<td>0.0582</td>
<td>0.8790</td>
<td>26.5217</td>
<td>27.4549</td>
</tr>
</tbody>
</table>
1. Video for which we would like to define a background image
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.

Copyright Piérard Sébastien, 2012
35 human experts participated to the study.

Unable to choose between LaBGen and LaBGen-P for 38 sequences.

LaBGen-P was preferred 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
Results for SBMnet 2016

<table>
<thead>
<tr>
<th>Method</th>
<th>Overall Average ranking</th>
<th>Average ranking across categories</th>
<th>Average AGE</th>
<th>Average pEPs</th>
<th>Average pCEPS</th>
<th>Average MSSSIM</th>
<th>Average PSNR</th>
<th>CQM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSCL [15]</td>
<td>1.17</td>
<td>4.88</td>
<td>5.9547</td>
<td>0.0524</td>
<td>0.0171</td>
<td>0.9410</td>
<td>30.8952</td>
<td>31.7049</td>
</tr>
<tr>
<td>LaBGen-OF [26]</td>
<td>2.17</td>
<td>4.56</td>
<td>8.1897</td>
<td>0.0566</td>
<td>0.0232</td>
<td>0.9412</td>
<td>29.8057</td>
<td>30.7066</td>
</tr>
<tr>
<td>BEWIS [24]</td>
<td>4.33</td>
<td>5.68</td>
<td>6.7094</td>
<td>0.0532</td>
<td>0.0286</td>
<td>0.9262</td>
<td>23.7728</td>
<td>26.8432</td>
</tr>
<tr>
<td>LaBGen [6]</td>
<td>4.83</td>
<td>7.75</td>
<td>6.7090</td>
<td>0.0631</td>
<td>0.0235</td>
<td>0.9260</td>
<td>29.0396</td>
<td>29.4668</td>
</tr>
<tr>
<td>NExtBI [26]</td>
<td>6.00</td>
<td>11.38</td>
<td>6.7776</td>
<td>0.0671</td>
<td>0.0227</td>
<td>0.9196</td>
<td>27.9944</td>
<td>28.8810</td>
</tr>
<tr>
<td>LaBGen-P [7]</td>
<td>6.33</td>
<td>8.50</td>
<td>7.0738</td>
<td>0.0706</td>
<td>0.0319</td>
<td>0.9278</td>
<td>26.4650</td>
<td>29.3195</td>
</tr>
<tr>
<td>Photomontage [3]</td>
<td>7.17</td>
<td>11.13</td>
<td>7.1950</td>
<td>0.0686</td>
<td>0.0257</td>
<td>0.9169</td>
<td>28.0113</td>
<td>28.8710</td>
</tr>
<tr>
<td>SC-SOBS-C4 [5]</td>
<td>8.33</td>
<td>9.63</td>
<td>7.5183</td>
<td>0.0711</td>
<td>0.0242</td>
<td>0.9160</td>
<td>27.6533</td>
<td>26.5601</td>
</tr>
<tr>
<td>MAGPRC [10]</td>
<td>9.50</td>
<td>9.68</td>
<td>8.3132</td>
<td>0.0654</td>
<td>0.0597</td>
<td>0.9401</td>
<td>29.4556</td>
<td>28.3152</td>
</tr>
<tr>
<td>Temporal median filter [2]</td>
<td>11.67</td>
<td>9.13</td>
<td>8.2761</td>
<td>0.0604</td>
<td>0.0546</td>
<td>0.9130</td>
<td>27.5304</td>
<td>26.4434</td>
</tr>
<tr>
<td>BE-AAPSA [14]</td>
<td>11.67</td>
<td>12.13</td>
<td>7.9066</td>
<td>0.0675</td>
<td>0.0447</td>
<td>0.9127</td>
<td>27.0714</td>
<td>27.9211</td>
</tr>
<tr>
<td>Bidirectional Analysis [13]</td>
<td>12.00</td>
<td>11.13</td>
<td>8.3449</td>
<td>0.0756</td>
<td>0.0181</td>
<td>0.9085</td>
<td>25.1722</td>
<td>27.1637</td>
</tr>
<tr>
<td>Bidirectional Analysis and Consensus Voting [12]</td>
<td>13.50</td>
<td>13.00</td>
<td>8.5816</td>
<td>0.0724</td>
<td>0.0257</td>
<td>0.9078</td>
<td>26.1018</td>
<td>27.1000</td>
</tr>
</tbody>
</table>

Click on method name for more details.
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
{BLauugraud, 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 SRMnet dataset by LaBGen and LaBGen-P with the parameters discussed in Section II C. The discontinuities observable between different spatial areas make the estimation visually incoherent. They are avoided with the LaBGen-P method.

Fig. 1. Background estimated for the Hoard sequence of the SRMnet dataset by LaBGen and LaBGen-P with the parameters discussed in Section II 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
URL: https://youtu.be/lcXHM42EeZo
Unsolved Issues

- Intermittent motion $\rightarrow$ **memoryless** motion detection **inefficient**.

- Strong illum. changes $\rightarrow$ **no intrinsic mechanism** for temporal consistency.
Conclusion
- **A lot of challenges** make background initialization difficult to solve.

- Methods based on various math. tools → **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 **evaluation methodology**.
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 :-)!!*
Thank you
References
<table>
<thead>
<tr>
<th>Reference</th>
<th>Title</th>
</tr>
</thead>
</table>
References


References


References


References


