



## CDnet 2014: An Expanded Change Detection Benchmark Dataset

Yi Wang, Pierre-Marc Jodoin, Fatih Porikli, Janusz Konrad, Yannick Benezeth, Prakash Ishwar

### ► To cite this version:

Yi Wang, Pierre-Marc Jodoin, Fatih Porikli, Janusz Konrad, Yannick Benezeth, et al.. CDnet 2014: An Expanded Change Detection Benchmark Dataset. IEEE CVPR Change Detection workshop, Jun 2014, United States. 8 p. hal-01018757

**HAL Id: hal-01018757**

**<https://u-bourgogne.hal.science/hal-01018757>**

Submitted on 4 Jul 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# CDnet 2014: An Expanded Change Detection Benchmark Dataset

Yi Wang<sup>1</sup> Pierre-Marc Jodoin<sup>1</sup> Fatih Porikli<sup>2</sup> Janusz Konrad<sup>4</sup> Yannick Benezeth<sup>3</sup> Prakash Ishwar<sup>4</sup>

<sup>1</sup> Université de Sherbrooke CREI Sherbrooke, J1K 2R1, Canada	<sup>2</sup> Australian National University NICTA Canberra ACT 2601, Australia	<sup>3</sup> Université de Bourgogne Le2i 21078 Dijon, France	<sup>4</sup> Boston University ECE Department Boston MA, 02215, USA
--	--	---	---

## Abstract

*Change detection is one of the most important low-level tasks in video analytics. In 2012, we introduced the [changedetection.net](http://changedetection.net) (CDnet) benchmark, a video dataset devoted to the evaluation of change and motion detection approaches. Here, we present the latest release of the CDnet dataset, which includes 22 additional videos (~70,000 pixel-wise annotated frames) spanning 5 new categories that incorporate challenges encountered in many surveillance settings. We describe these categories in detail and provide an overview of the results of more than a dozen methods submitted to the IEEE Change Detection Workshop 2014. We highlight strengths and weaknesses of these methods and identify remaining issues in change detection.*

## 1. Introduction

Change (motion) detection is among the most frequently used preprocessing tasks in a wide variety of applications such as people counting, traffic video analysis, object tracking, mobility monitoring and video retrieval to name just a few. Most change detection methods attribute a “changed” label to a pixel whose photometric properties deviate from those of the background scene.

Even though many approaches perform well when the camera is static, the illumination is sufficient, the weather is favorable and the image is artifact-free, it is widely documented that change detection methods are error-prone whenever one of the underlying assumptions is violated [3, 4, 6, 7, 9]. Up until recently, there was no widely accepted dataset that researchers could rely on to compare their methodology. The lack of a common benchmark dataset made it difficult to identify methods that are most effective in dealing with more general and challenging scenarios. Instead, past work often tended to “specialize” on a particular question, e.g., background motion, unstable lighting, and shadows. Moreover, authors often opted to compare their ideas against easily-implementable (and in many cases outdated) change detection approaches. This caused an overwhelming attention to be given to a limited number of outdated methods while marginalizing the more competent recent ones. Since the performance of a method would vary from one implementation to another, especially con-

sidering different parametrizations and post-processing options, it has been very problematic to reproduce the same results as the ones given in the original paper. The fact that past work often reported results on nonpublic sequences made a fair comparison an elusive goal.

To respond to this need, the first version of the [changedetection.net](http://changedetection.net) (CDnet) dataset was released in 2012. The goal was to provide a balanced dataset depicting various scenarios that are common in change detection. For this, we prepared a dataset of 31 videos (~90,000 manually annotated frames) categorized into 6 challenges. Since each category is associated with a specific change detection problem, e.g., dynamic background, shadows, etc., CDnet enabled an objective identification and ranking of methods that are most suitable for a specific problem as well as competent overall. In addition, we provided an online evaluation facility to help researchers compare their algorithms with the state-of-the-art including evaluation tools, pixel-accurate ground-truth frames and online ranking tables. According to Google Analytics, the CDnet website was visited by more than 12,000 individual users during the last 24 months and results from 34 different methods were uploaded to our system (see the 2012 results section on the CDnet website). We received very positive comments and several suggestions for further improvement.

Given this feedback, we prepared a second version of the dataset: the 2014 CDnet. It includes 22 additional videos in 5 new categories. In the remainder of the paper, we provide an overview of the existing change detection datasets as well as survey papers, and summarize the 2012 CDnet dataset. Then, we present the 2014 CDnet dataset. We discuss its categories, ground-truth annotations, performance metrics, and a summary of the comparative rankings of the methods that competed in the Change Detection Workshop (CDW) held in conjunction with CVPR 2014.

## 2. Previous Work

### 2.1. Datasets and Surveys

Several datasets have been used in the past to evaluate change detection algorithms. We list 18 of those datasets in Table 1. Of these datasets, seven have been initially developed to validate tracking and pattern recognition methods

Dataset	Description	Ground truth
Wallflower [27]	7 short video clips, each representing a specific challenge such as illumination change, background motion, etc	Pixel-based labeling of one frame per video.
PETS [31]	Many videos aimed at evaluating tracking algorithms	Bounding boxes.
CAVIAR	80 staged indoor videos representing different human behaviors such as walking, browsing, shopping, fighting, etc.	Bounding boxes.
i-LIDS	Long videos meant for action recognition. Shows parked vehicles, abandoned objects, walking people, doorways	Not fully labeled.
ETISEO	More than 80 videos meant to evaluate tracking and event detection methods.	High-level labels such as bounding box, object class, event type.
ViSOR 2009 [29]	Web archive with 500 short videos (most less than 10 seconds)	Bounding boxes.
BEHAVE 2007	7 videos shot by the same camera showing human interactions such as walking in group, meeting, splitting, etc.	Bounding boxes.
VSSN 2006	9 semi-synthetic videos composed of a real background and artificially-moving objects. The videos contain animated background, illumination changes and shadows.	Pixel-based labeling of each frame.
IBM	15 videos taken from PETS 2001 plus additional videos.	Bounding box around each moving object in 1 frame out of 30.
Karlsruhe	4 grayscale traffic videos under various weather conditions.	10 frames/video Pixel-wise labels.
Li <i>et al.</i> [12]	10 short videos with illumination changes and dynamic back.	10 frames/video Pixel-wise labels.
Karaman [11]	5 videos from different sources (web, “art live” project, etc.) with various illuminations and compression artifacts	Pixel-wise labels of each frame.
cVSG [26]	15 Semi-synthetic videos with various levels of textural complexity, background motion, moving object speed, size and interaction.	Pixel-wise labeling by filming moving objects in front of a blue-screen pasted on background videos.
LIMU	8 simple videos, some borrowed from PETS2001.	Pixel-wise labeling, 1 out of 15 frames.
UCSD	18 short videos with background motion and/or camera motion.	Pixel-wise labeling.
SZTAKI	5 indoor/outdoor videos with shadows	Pixel-wise labeling of foreground, shadows for a subset of frames .
BMC 2012 [28]	29 outdoor videos, most being synthetic.	Pixel-wise labeling for 10 synthetic and 9 real videos. Labeling of real videos is for a small subset of frames.
Brutzer <i>et al.</i> [7]	Computer-generated videos showing a street corner. Includes illumination changes, dynamic background, shadows, noise.	Pixel-wise labeling.

Table 1. Overview of 18 video databases often used to validate change detection methods.

(namely PETS, CAVIAR, i-LIDS, ETISEO, ViSOR 2009, BEHAVE 2007, IBM). Although these datasets are challenging for these applications, they only contain day-time videos with fixed background, constant illumination, few shadows and no camera jitter. Other datasets contain synthetic or semi-synthetic videos (VSSN 2006, cVSG, BMC 2012, and Brutzer *et al.*) whose content is not always realistic. The remaining datasets either contain a limited number of videos or focus on one specific type of action, e.g., indoor human actions or motor traffic scenes. As a consequence, it is difficult to evaluate how robust motion detection methods are when looking at the results reported on these datasets. More details on these datasets can be found in [9].

A number of survey papers have been written on the topic of motion detection (Table 2). Note that, some of these surveys use the benchmark videos, while others use their own, non-public datasets. Although they provide a good overview of existing methods, the statistics reported in these surveys have not been computed on a general, unbiased, balanced dataset composed of real (camera-captured) videos. Most of these studies report fairly simple motion detection methods and do not discuss the performance of recent (and more complex) solutions.

Let us mention the 2014 handbook by Bouwmans *et al* [6] which, to our knowledge, is the most complete manuscript devoted to change detection recently published.

Survey	Description and Benchmark
Bouwmans <i>et al.</i> , 2011 [4]	Potentially the most complete survey to date with more than 350 references. The paper reviews methods spanning 6 motion detection categories and the features used by each method. It provides benchmarking results on the Wallflower dataset.
Benezeth <i>et al.</i> , 2010 [3]	Survey presenting benchmarking results obtained on 29 videos (15 camera-captured, 10 semi-synthetic, and 4 synthetic) taken from PETS 2001, the IBM dataset, and the VSSN 2006 dataset.
Nascimento-Marques, 2006 [17]	Contains benchmarks obtained on a single PETS 2001 video with pixel-based labeling.
Prati <i>et al.</i> , 2001 [21]	Contains benchmarks of indoor sequences containing one moving person. 112 frames were labeled.
Parks and Fels, 2008 [19]	Results for 7 motion detection methods and evaluation of post-processing on their performance. They used 7 outdoor and 6 indoor videos containing different challenges such as dynamic backgrounds, shadows and various lighting conditions.
Bashir and Porikli, 2006 [2]	Performance evaluation of tracking algorithms using the PETS 2001 dataset by comparing the detected bounding box locations with the ground-truth.
Bouwmans <i>et al.</i> , 2008 [5]	Survey of GMM methods. Provides benchmarking results on the Wallflower dataset.
Radke <i>et al.</i> , 2005 [23]	Extensive survey of several motion detection methods. Most of the discussion in the paper was related to background subtraction methods, pre- and post-processing, and methodologies to evaluate performances. Contains no quantitative evaluation.
Piccardi, 2004 [20]	Review of 7 background subtraction methods which highlights their strengths and weaknesses. Contains no quantitative evaluation.
Rosin-Ioannidis, 2003 [24]	Report results for 8 method with home-made videos showing balls rolling on the floor.
Brutzer <i>et al.</i> , 2011 [7]	Survey which reports benchmarking results for 8 motion detection method on the computer-generated Brutzer 2011 dataset.

Table 2. 11 change detection survey papers.

## 2.2. CDnet 2012 Dataset

The 2012 dataset was composed of nearly 90,000 frames in 31 video sequences representing various challenges divided into 6 categories ( Fig. 1):

1. **Baseline** contains 4 videos with a mixture of mild challenges of the next 4 categories. These videos are fairly easy and are provided mainly as reference.
2. **Dynamic Background** contains 6 videos depicting outdoor scenes with strong background motion.
3. **Camera Jitter** represents 4 videos captured with unstable cameras.
4. **Shadow** is composed of 6 videos with both strong and soft moving and cast shadows.
5. **Intermittent Object Motion** contains 6 videos with scenarios known for causing ghosting artifacts (*e.g.* contains still objects that suddenly start moving).
6. **Thermal** is composed of five videos captured by far-infrared cameras.

The results of the CDW-2012 workshop brought to light a number of interesting findings and helped pinpoint critical unresolved issues in motion and change detection. Our key findings from the 2012 workshop include the following:

1. Contrary to popular belief, videos with small recurrent background motion (ripples on the water, trees shaken by the wind) do not pose a significant challenge (any more). This conclusion applies to baseline videos.
  2. None of the methods tested is robust to hard shadows, ghosting artifacts, intermittent motion, and camouflage. These are open issues that are yet to be solved.
  3. Contrary to common belief, detecting humans and moving objects in thermal videos is not trivial. It is often accompanied by heat reflections and camouflage effects that no method handles well.
  4. An interesting finding is that methods appear to be complementary in nature: the best-performing methods can be beaten by combining several of them with a majority vote. This suggests that future research should consider modular change detection methods that incorporate a combination of several complementary change detection strategies.
  5. The f-measure correlates most strongly with the average rankings produced by our evaluation algorithm.
- Detailed findings of CDW-2012 can be found in [9].

## 3. Extended CDnet 2014 Dataset

Guided by the feedback from other researchers, we prepared a new set of videos representing 5 additional cate-

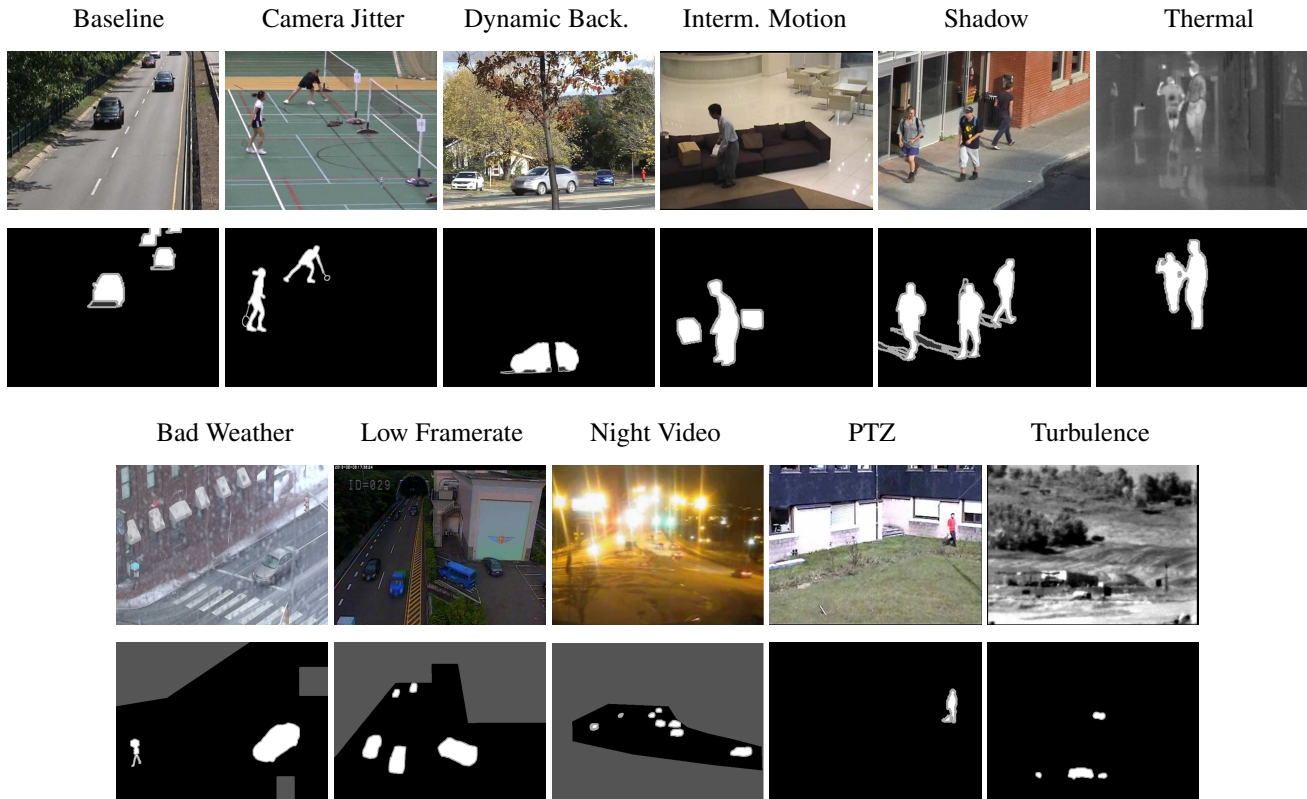


Figure 1. Sample video frames from each of the 11 categories in the new dataset available at [www.changedetection.net](http://www.changedetection.net). The categories in the first row exist in both 2012 and 2014 dataset while the ones in the second row only exist in 2014 dataset.

gories incorporating challenges not addressed in the 2012 CDnet dataset. In total, more than 70,000 frames have been captured, and then manually segmented and annotated by a team of 13 researchers from 7 universities.

The 2014 CDnet dataset provides realistic, camera-captured (without CGI), diverse set of indoor and outdoor videos like the 2012 CDnet. These videos have been recorded using cameras ranging from low-resolution IP cameras, higher resolution consumer grade camcorders, commercial PTZ cameras to near-infrared cameras. As a consequence, spatial resolutions of the videos in the 2014 CDnet vary from  $320 \times 240$  to  $720 \times 486$ . Due to the diverse lighting conditions present and compression parameters used, the level of noise and compression artifacts significantly varies from one video to another. Duration of the videos are from 900 to 7,000 frames. Videos acquired by low-resolution IP cameras suffer from noticeable radial distortion. Different cameras have different hue bias due to different white balancing algorithms employed. Some cameras apply automatic exposure adjustment resulting in global brightness fluctuations in time. Frame rate also varies from one video to another, often as a result of limited bandwidth. Since these videos have been captured under a wide range of settings, the extended 2014 CDnet dataset does not favour a

certain family of change detection methods over others.

### 3.1. Video Categories

There are 5 new categories as shown in Fig. 1. Similarly to the 2012 dataset, the change detection challenge in a category is unique to that category. Such a grouping is essential for an unbiased and clear identification of the strengths and weaknesses of different methods. These categories are:

**Challenging Weather:** This category contains 4 outdoor videos showing low-visibility winter storm conditions. This includes two traffic scenes in a blizzard, cars and pedestrians at the corner of a street and people skating in the snow. These videos present a double challenge: in addition to snow accumulation, the dark tire tracks left in the snow have potential to cause false positives.

**Low Frame-Rate:** In this category 4 videos, all recorded with IP cameras, are included. The frame rate varies from 0.17 fps to 1 fps due to limited transmission bandwidth. By nature, these videos show “erratic motion patterns” of moving objects that are hard (if not impossible) to correlate. Optical flow might be ineffective for these videos. One sequence is particularly challenging (*port\_0\_17fps*), which shows boats and people coming in and out of a harbour, as



the low framerate accentuates the wavy motion of moored boats causing false detections.

**Night:** This category has 6 motor traffic videos. The main challenge is to cope with low-visibility of vehicles yet their very strong headlights that cause over saturation. Headlights cause halos and reflections on the street.

**PTZ:** We included 4 videos in this category: one video with a slow continuous camera pan, one video with an intermittent pan, one video with a 2-position patrol-mode PTZ, and one video with zoom-in/zoom-out. The PTZ category by itself requires different type of change detection techniques in comparison to static camera videos.

**Air Turbulence:** This category contains 4 videos showing moving objects depicted by a near-infrared camera at noon during a hot summer day. Since the scene is filmed at a distance (5 to 15 km) with a telephoto lens, the heat causes constant air turbulence and distortion in frames. This results in false positives. The size of the moving objects also varies significantly from one video to another. The air turbulence category presents very similar challenges to those arising in long-distance remote surveillance applications.

### 3.2. Ground-Truth Labels

For consistency, we use the same labels as those in the 2012 CDnet dataset. Each frame has been manually annotated at pixel level, with the following 5 labels:

1. **Static** pixels are assigned grayscale value of 0.
2. **Shadow** pixels are assigned grayscale value of 50. The *Shadow* label is associated with hard and well-defined moving shadows such as the one in Fig. 2.
3. **Non-ROI**<sup>1</sup> pixels (*i.e.* outside of the ROI) are assigned grayscale value of 85. The first few hundred frames of each video sequence are labelled as *Non-ROI* to prevent the corruption of evaluation metrics due to errors during initialization. The *Non-ROI* label prevents also the metrics from being corrupted by activities unrelated to the category considered.
4. **Unknown** grayscale value of 170 assigned to pixels that are half-occluded or corrupted by motion blur.
5. **Moving** pixels are assigned grayscale value of 255.

Please note that the evaluation metrics discussed in Section 3.3 consider the *Shadow* pixels as *Static* pixels.

### 3.3. Evaluation Metrics

As in 2012, results are presented by reporting the average performance of each method, for each video category with respect to 7 different performance metrics [9]. Let  $TP$  be the number of true positives,  $TN$  the number of true negatives,  $FN$  the number of false negatives, and  $FP$  the

<sup>1</sup>ROI stands for Region Of Interest.



Figure 2. Sample frames showing 5-class ground-truth label fields.

number of false positives. For each method and each video category, the following 7 metrics are computed:

1. Recall (Re):  $TP/(TP+FN)$
2. Specificity (Sp):  $TN/(TN+FP)$
3. False Positive Rate (FPR):  $FP/(FP+TN)$
4. False Negative Rate (FNR):  $FN/(TN+FP)$
5. Percentage of Wrong Classifications (PWC):  $100(FN+FP)/(TP+FN+FP+TN)$
6. Precision (Pr):  $TP/(TP+FP)$
7. *F*-measure (or *F1* score):  $2 \frac{Pr \cdot Re}{Pr+Re}$

For the *Shadow* category, we also provide an average FPR that is confined to the hard-shadow areas (FPR-S).

In order to easily assess the various change detection methods, these metrics are then combined into two metrics  $R$  and  $RC$  [9].  $R$  represents an average ranking computed across all overall-average metrics.  $RC$  is an average ranking computed across all categories. Results with  $R$  and  $RC$  are also presented in Table 3.

## 4. Methods Tested

A total of 14 change detection methods were evaluated for the IEEE Change Detection Workshop 2014 [1]. Of these, 3 are relatively simpler methods that rely on plain background subtraction and two of these use color features: the Euclidean and Mahalanobis distance methods as described in [3]. Two methods are older but frequently-cited: KDE-based estimation by Elgammal *et al.* [8] and GMM by Stauffer and Grimson [25]. We also include a K-nearest neighbor (KNN) method as well as a recursive GMM method with an improved update of the Gaussian parameters and an automatic selection of the number of modes [32].

Among the more recent methods is FTSG [22] which uses a three-step procedure consisting of 1) moving object detection with two complementary pixel-level motion detection methods based on the trace of a flux tensor and a

Method	RC	R	Re	Sp	FPR	FNR	PWC	F-Measure	Pr
Majority Vote-3	1.55	2.14	0.76	0.99	0.0054	0.24	1.31	0.75	0.81
FTSG [22]	1.82	2.14	0.77	0.99	0.0078	0.23	1.38	0.73	0.80
SuBSENSE [18]	2.36	2.71	0.81	0.99	0.0116	0.19	1.84	0.73	0.75
Majority Vote-all	4.00	7.29	0.66	0.99	0.0112	0.34	2.15	0.64	0.74
CwisarDH [10]	4.00	4.57	0.66	0.99	0.0052	0.34	1.53	0.68	0.77
Spectral-360 [16]	5.36	5.43	0.73	0.99	0.0139	0.27	2.27	0.67	0.71
Bin Wang Apr 2014 [30]	7.27	6.71	0.70	0.98	0.0206	0.30	2.90	0.66	0.72
KNN [32]	7.91	8.14	0.67	0.98	0.0198	0.34	3.32	0.59	0.68
SC-SOBS [15]	8.45	8.29	0.76	0.95	0.0453	0.24	5.15	0.60	0.61
KDE-ElGammal [8]	9.36	10.43	0.74	0.95	0.0481	0.26	5.63	0.57	0.58
Mahalanobis distance [3]	9.82	8.86	0.16	0.99	0.0069	0.84	3.48	0.23	0.74
GMM Stauffer-Grimson [25]	10.36	9.14	0.68	0.98	0.0250	0.32	3.77	0.57	0.60
CP3-online [13]	10.82	9.14	0.72	0.97	0.0295	0.28	3.43	0.58	0.56
GMM Zivkovic [32]	11.27	11.43	0.66	0.97	0.0275	0.34	4.00	0.56	0.60
Multiscale Spatio-Temp BG Model [14]	12.45	12.71	0.66	0.95	0.0458	0.34	5.55	0.51	0.55
Euclidean distance [3]	13.82	13.00	0.68	0.94	0.0551	0.32	6.54	0.52	0.55

Table 3. Overall results across all categories (RC: average ranking across categories, R: average overall ranking).

variant of the conventional GMM, 2) fusion motion detection results, and 3) removal of ghosting artifacts. In SuBSENSE [18], color and local binary similarity patterns are used to make pixel-level decisions with automatic adjustment of tuning parameters for locally adapting to changing input. CwisarDH [10] uses a neural network where a random collection of the previous values of a pixel constitutes a random access memory node and the responses of such nodes are aggregated to compute a discriminator response. Maddalena and Petrosino’s SC-SOBS method [15] also uses machine learning but with a self-organizing neural map. Spectral-360 [16] is based on the correlation between the diffuse spectral reflectance components of a new video frame and an evolving background model derived from recent training frames. The method of [30] compares the current pixel value with one long-term and several short-term adaptive templates that are discarded based on a measure of efficacy rather than age or random selection. For CP3-online [13], instead of modeling the background model for each pixel individually, they model the color distribution of pixels with strong spatial correlation. The authors argue that such spatial model is robust to sudden illumination changes. A multiresolution approach to background subtraction is developed in [14] using a 3-scale spatio-temporal color/luminance Gaussian pyramid background model and a system of pixel neighborhoods at each scale.

For each method, only one set of parameters was used for all the videos. These parameters were selected according to the authors’ recommendations or, when not available, were adjusted to enhance the overall results. All parameters are available on the CDW-2014 website.

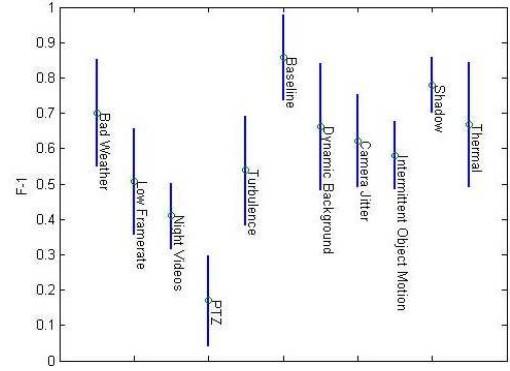


Figure 3. Mean and  $\sigma$  of the F-Measure over all methods.

## 5. Experimental Results

In order to give the reader an intuitive understanding of the overall performance of all 14 methods, we put in Fig. 3 the mean and standard deviation of the F-measure for all methods within each category. Without much surprise, the PTZ category has the lowest performance. The second most difficult category is the night videos, followed by the low framerate and turbulence. Surprisingly, every method performed relatively well on the bad weather videos. Overall, the performance obtained on the 2014 categories is lower than for the 2012 categories. We also report the median metrics obtained by all methods in the new 5 categories as shown in Table 4.

In Table 3, we report the overall results for all 14 methods, which we sorted according to their average ranking across categories (RC). We also report results obtained after combining all the methods (*Majority Vote-all*) and the top 3

Category	F-Measure	FPR	FNR	PWC
Bad Weather	0.74	0.0018	0.32	0.76
Low Framerate	0.54	0.0058	0.35	1.34
Night Videos.	0.41	0.0304	0.42	4.95
PTZ	0.14	0.1149	0.34	11.79
Turbulence	0.50	0.0069	0.25	0.82

Table 4. Median F-Measure, FPR, FNR and PWC obtained by all 14 methods for each category.

Category	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>
Baseline	SC-SOBS	SuBSENSE	CwisarDH
Dynamic	FTSG	Bin Wang	CwisarDH
Camera Jitter	CwisarDH	SuBSENSE	FTSG
Intermittent	FTSG	Bin Wang	SuBSENSE
Shadow	FTSG	SuBSENSE	CwisarDH
Thermal	FTSG	CwisarDH	KDE
Bad Weather	FTSG	SuBSENSE	Bin Wang
Low Framerate	FTSG	SuBSENSE	Spectral-360
Night Videos	SuBSENSE	FTSG	Spectral-360
PTZ	FTSG	SuBSENSE	CwisarDH
Turbulence	SuBSENSE	Bin Wang	FTSG

Table 5. Three best methods for each category

methods FTSG, SuBSENSE and CwisarDH (*Majority Vote-3*). We did so with a pixel-based majority vote. As one can see, even by combining basic methods, *Majority Vote-all* outperforms every method except FTSG and SuBSENSE. As for *Majority Vote-3*, it outperforms every other motion detection method. This is a strong indication that no single method decisively outperforms other ones. On the contrary, these methods seam complementary by nature. Table 5 underlines a similar conclusion since no method is systematically the best in each category.

In order to identify where the methods fail, we integrated the error at each pixel of each frame and for every method. This led to error maps similar to the ones shown in Fig 4. In these images, red, green, white and black stand for the false negative, false positive, true positive and true negative, respectively. Pixels with saturated red and green indicate that every method failed at those pixels. After analyzing these error maps, we came to identify the most glaring issues that no single method handles well:

1. **PTZ**: any camera motion (pan, tilt or zoom) causes major false positives.
2. **Night Videos**: the lack of illumination causes numerous false negatives while headlight reflections cause systematic false positives.
3. **Shadow**: Hard shadows are still a challenge for every method.
4. **Intermittent Object Motion**: Any object which stops moving for some time, eventually ends up being mis-

detected. A similar situation occurs when a background object is removed from the scene.

5. **Turbulence**: Air turbulence causes the systematic occurrence of false positives.

We also compare the results for the 6 categories which are included in both the 2012 and 2014 dataset. For each category, we took the average performance of the top 4 methods reported on the 2012 dataset and the 2014 dataset (see the results section on the CDnet website). By doing this, we wanted to see if the 2012 methods which had to process only these 6 categories were doing better than those that had to account for 11 categories. As shown in Table 6, except for the camera jitter category for which the 2012 methods are more accurate, the results obtained by the 2012 and 2014 methods are very similar.

## Acknowledgements

We thank Shaozi Li, Weiyan Zhuang, Yaping You, Lucas Liang, and Luke Sorenson for manual ground-truthing, Marc Vandroogenbroeck for providing a PTZ video, Oognet.pl for providing free access to their surveillance cameras, and Dotan Asselmann for providing turbulence videos and helping with manual ground-truthing.

## References

- [1] 2nd IEEE Change Detection Workshop, 2014, in conjunction with CVPR. [www.changedetection.net](http://www.changedetection.net). 5
- [2] F. Bashir and F. Porikli. Performance evaluation of object detection and tracking systems. In *Proc. IEEE Int. Workshop on Performance Evaluation of Tracking Systems*, 2006. 3
- [3] Y. Benezeth, P.-M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger. Comparative study of background subtraction algorithms. *J. of Elec. Imaging*, 19(3):1–12, 2010. 1, 3, 5, 6
- [4] T. Bouwmans. Recent advanced statistical background modeling for foreground detection: A systematic survey. *Recent Patents on Computer Science*, 4(3), 2011. 1, 3
- [5] T. Bouwmans, F. E. Baf, and B. Vachon. Background modeling using mixture of gaussians for foreground detection: A survey. *Recent Patents on Computer Science*, 1(3):219–237, 2008. 3
- [6] T. Bouwmans, F. Porikli, B. Hferlin, and A. Vacavant. *Background Modeling and Foreground Detection for Video Surveillance*. Chapman and Hall/CRC, 2014. 1, 2
- [7] S. Brutzer, B. Hoferlin, and G. Heidemann. Evaluation of background subtraction techniques for video surveillance. In *Proc. IEEE Conf. Computer Vision Pat. Recog.*, pages 1937–1944, 2011. 1, 2, 3
- [8] A. Elgammal, R. Duraiswami, D. Harwood, and L. Davis. Background and foreground modeling using nonparametric kernel density for visual surveillance. *Proc. IEEE*, 90:1151–1163, 2002. 5, 6
- [9] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. Changedetection.net: A new change detection benchmark dataset. In *IEEE CVPR change detection workshop*, 2012. 1, 2, 3, 5
- [10] M. D. Gregorio and M. Giordano. Change detection with weightless neural networks. In *IEEE Workshop on Change Detection*, 2014. 6
- [11] M. Karaman, L. Goldmann, D. Yu, and T. Sikora. Comparison of static background segmentation methods. In *Proc. SPIE Visual Communications and Image Process.*, pages 2140–2151, 2005. 2
- [12] L. Li, W. Huang, I. G. Yu-Hua, and Q. Tian. Statistical modeling of complex backgrounds for foreground object detection. *IEEE Trans. Image Process.*, 13(11):1459–1472, 2004. 2
- [13] D. Liang and S. Kaneko. Improvements and experiments of a compact statistical background model, 2014. 6



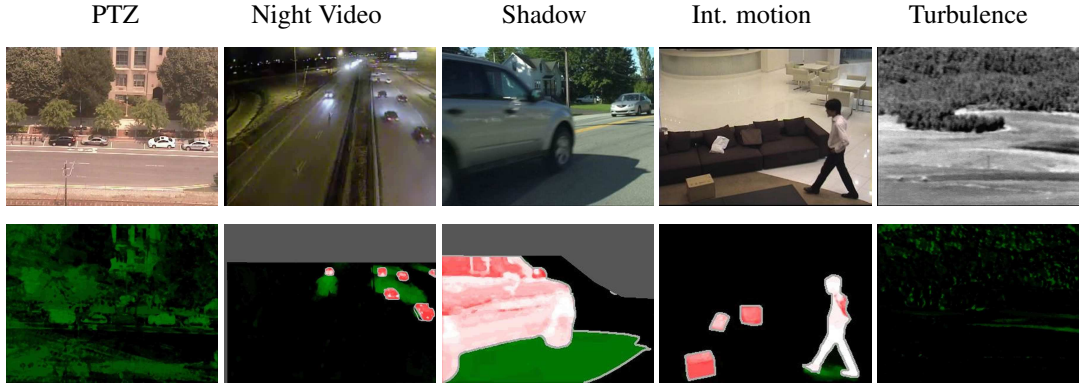


Figure 4. Error maps showing systematic errors. Red: false positives, green: false negatives, white: true positives and black: true negatives.

Category	Year	Re	Sp	FPR	FNR	PWC	F-Measure	Pr
Baseline	2012	0.92	1.00	0.0021	0.08	0.42	0.93	0.94
	2014	0.94	1.00	0.0022	0.06	0.45	0.93	0.93
Dynamic Background	2012	0.87	1.00	0.0017	0.13	0.28	0.84	0.84
	2014	0.85	1.00	0.0018	0.15	0.35	0.84	0.86
Camera Jitter	2012	0.79	0.99	0.0090	0.21	1.65	0.80	0.85
	2014	0.77	0.99	0.0132	0.23	2.12	0.75	0.76
Intermittent Motion	2012	0.74	0.98	0.0185	0.26	3.24	0.69	0.75
	2014	0.69	0.99	0.0089	0.31	3.21	0.68	0.79
Shadow	2012	0.92	0.99	0.0091	0.08	1.19	0.87	0.84
	2014	0.91	0.99	0.0092	0.09	1.26	0.87	0.84
Thermal	2012	0.69	1.00	0.0028	0.31	1.86	0.75	0.93
	2014	0.71	1.00	0.0049	0.29	1.53	0.77	0.90

Table 6. Comparison of methods tested in CDW-2012 and CDW-2014 on the original categories. For each category, the top 4 methods are used to represent its performance.

- [14] X. Lu. A multiscale spatio-temporal background model for motion detection. In *Proc. IEEE Int. Conf. Image Processing*, 2014. 6
- [15] L. Maddalena and A. Petrosino. The SOBS algorithm: what are the limits? In *IEEE Workshop on Change Detection*, 2012. 6
- [16] M. M. Sedky and C. C. Chibelushi. Spectral-360: A physical-based technique for change detection. In *IEEE Workshop on Change Detection*, 2014. 6
- [17] J. Nascimento and J. Marques. Performance evaluation of object detection algorithms for video surveillance. *IEEE Trans. Multimedia*, 8(8):761–774, 2006. 3
- [18] G.-A. B. P.-L. St-Charles and R. Bergevin. Flexible background subtraction with self-balanced local sensitivity. In *IEEE Workshop on Change Detection*, 2014. 6
- [19] D. Parks and S. Fels. Evaluation of background subtraction algorithms with post-processing. In *Proc. IEEE Int. Conf. on Advanced Video and Signal-based Surveillance*, pages 192–199, 2008. 3
- [20] M. Piccardi. Background subtraction techniques: a review. pages 3099–3104, 2004. 3
- [21] A. Prati, R. Cucchiara, I. Mikic, and M. Trivedi. Analysis and detection of shadows in video streams: A comparative evaluation. In *Proc. IEEE Conf. Computer Vision Pat. Recog.*, pages 571–577, 2001. 3
- [22] G. S. R. Wang, F. Bunyak and K. Palaniappan. Static and moving object detection using flux tensor with split gaussian models. In *IEEE Workshop on Change Detection*, 2014. 5, 6
- [23] R. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms: A systematic survey. *IEEE Trans. Image Process.*, 14:294–307, 2005. 3
- [24] P. Rosin and E. Ioannidis. Evaluation of global image thresholding for change detection. *Pattern Recog. Lett.*, 24:2345–2356, 2003. 3
- [25] C. Stauffer and E. Grimson. Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(8):747–757, 2000. 5, 6
- [26] F. Tiburzi, M. Escudero, J. Bescos, and J. Martinez. A ground truth for motion-based video-object segmentation. In *Proc. IEEE Int. Conf. Image Processing*, pages 17–20, 2008. 2
- [27] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *Proc. IEEE Int. Conf. Computer Vision*, volume 1, pages 255–261, 1999. 2
- [28] A. Vacavant, T. Chateau, A. Wilhelm, and L. Lequievre. A benchmark dataset for outdoor foreground/background extraction. In *ACCV Workshops*, pages 291–300, 2012. 2
- [29] R. Vezzani and R. Cucchiara. Video surveillance online repository (visor): an integrated framework. *Multimedia Tools and Applications*, 50(2):359–380, 2010. 2
- [30] B. Wang and P. Dudek. A fast self-tuning background subtraction algorithm. In *IEEE Workshop on Change Detection*, 2014. 6
- [31] D. Young and J. Ferryman. PETS metrics: Online performance evaluation service. In *Proc. IEEE Int. Workshop on Performance Evaluation of Tracking Systems*, pages 317–324, 2005. 2
- [32] Z. Zivkovic and F. V. D. Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recog. Lett.*, 27:773–780, 2006. 5, 6