changedetection.net: A New Change Detection Benchmark Dataset

Nil Goyette[1] Pierre-Marc Jodoin[1] Fatih Porikli[2] Janusz Konrad[3] Prakash Ishwar[3]

[1]Université de Sherbrooke
MOIVRE
Sherbrooke, J1K 2R1, Canada

[2]Mitsubishi Electric Research Laboratories
Imaging Group
Cambridge MA, 02139, USA

[3]Boston University
ECE Department
Boston MA, 02215, USA

Abstract

Change detection is one of the most commonly encountered low-level tasks in computer vision and video processing. A plethora of algorithms have been developed to date, yet no widely accepted, realistic, large-scale video dataset exists for benchmarking different methods. Presented here is a unique change detection benchmark dataset consisting of nearly 90,000 frames in 31 video sequences representing 6 categories selected to cover a wide range of challenges in 2 modalities (color and thermal IR). A distinguishing characteristic of this dataset is that each frame is meticulously annotated for ground-truth foreground, background, and shadow area boundaries – an effort that goes much beyond a simple binary label denoting the presence of change. This enables objective and precise quantitative comparison and ranking of change detection algorithms. This paper presents and discusses various aspects of the new dataset, quantitative performance metrics used, and comparative results for over a dozen previous and new change detection algorithms. The dataset, evaluation tools, and algorithm rankings are available to the public on a website[1] and will be updated with feedback from academia and industry in the future.

1. Introduction

Detection of change, and in particular motion, is a fundamental low-level task in many computer vision and video processing applications. Examples include visual surveillance (people counting, crowd monitoring, action recognition, anomaly detection, forensic retrieval, etc.), smart environments (occupancy analysis, parking lot management, etc.), and content retrieval (video annotation, event detection, object tracking). Change detection is closely coupled with higher level inference tasks such as detection, localization, tracking, and classification of moving objects, and is often considered to be critical preprocessing step. Its importance can be gauged by the large number of algorithms that have been developed to-date and the even larger number of articles that have been published on this topic. A quick search for 'motion detection' on IEEE Xplore© returns over 4,400 papers.

Among the many variants of change detection algorithms, there seems to be no single algorithm that competently addresses all of the inherent real-life challenges including sudden illumination variations, background movements, shadows, camouflage effects (photometric similarity of object and background) and ghosting artifacts (delayed detection of a moving object after it has moved away). Furthermore, due to the tremendous effort required to generate a benchmark dataset that contains pixel precision ground-truth labels and provides a balanced coverage of the range of challenges representative of the real world, no comprehensive large-scale evaluation of change detection has been conducted to date.

The lack of a comprehensive dataset has a number of negative implications. Firstly, it makes it difficult to ascertain with confidence which algorithms would perform robustly when the assumptions they are built upon are violated. Moreover, many algorithms tend to *overfit* specific scenarios. For example, a method may be tuned to be robust to shadows but may not be as robust to background motion. A dataset that includes many different scenarios and uses a variety of performance measures would go a long way towards providing an objective assessment. Secondly, not all authors are willing to (or have the resources to) compare their methods against the most advanced and promising approaches. As a consequence, an overwhelming importance has been accorded to a small subset of easily implementable methods such as [23, 9, 26] that were developed in the late 1990's. The more recent and advanced methods have been marginalized as a result. Besides, the implementation of the same method varies significantly from one research group to another in the choice of parameters and the use of other pre- and post-processing steps. Thirdly, the fact that authors often use their own data (that are not widely available to everyone) makes a fair comparison much more problematic if not impossible.

Recognizing the importance of change detection to the computer vision and video processing communities, we have prepared a unique change detection benchmark dataset: changedetection.net (CDnet) that consists of nearly

[1] www.changedetection.net

978-1-4673-1612-5/12/$31.00 ©2012 IEEE

90,000 frames in 31 video sequences representing 6 video categories (including thermal). This new dataset is the foundation of the 2012 IEEE Change Detection Workshop [1]. CDnet contains diverse motion and change detection challenges in addition to typical indoor and outdoor scenes that are encountered in most surveillance, smart environments, and video analytics applications. A distinguishing feature of CDnet is the fact that each image is meticulously annotated for ground-truth foreground, background, and shadow region boundaries; an effort that goes much beyond a simple binary label denoting the presence of the change. The existence of ground-truth masks permits a precise comparison and ranking of change detection algorithms. CDnet also supplies a selection of evaluation tools in *MATLAB* and *Python* for quantitatively assessing the performance of different methods according to 7 distinct metrics.

The overarching objectives of CDnet and its associated workshop can be listed as:

1. To provide the research community with a rigorous and comprehensive scientific benchmarking facility, a rich dataset of videos, a set of utilities, and an access to author-approved algorithm implementations for testing and ranking of existing and new algorithms for motion and change detection. The already extensive dataset will be regularly revised and expanded with feedback from the academia and industry.

2. To establish, maintain, and update a rank list of the most accurate motion and change detection algorithms in the various categories for years to come.

3. To help identify the remaining challenges in order to provide focus for future research.

Next, we provide an overview of the existing datasets and then present the details of CDnet including its categories, ground-truth annotations, performance metrics, and a summary of the comparative rankings of the methods that we tested at the IEEE Change Detection Workshop held in conjunction with CVPR 2012.

2. Overview of Prior Efforts

Several datasets and survey papers have been presented for the evaluation of change detection algorithms in the past.

2.1. Previous Datasets

Without aiming to be exhaustive, we list below a few key datasets and describe their characteristics:

- Wallflower [25]: This is a fairly well-known dataset that continues to be used today. It contains 7 videos, each representing a specific challenge such as illumination change, background motion, etc. Only one frame per video has been labeled.

- PETS [27]: The Performance Evaluation of Tracking and Surveillance (PETS) program was launched with the goal of evaluating visual tracking and surveillance algorithms. The program has been collecting videos for the scientific community since the year 2000 and now contains several dozen videos. Many of these videos have been manually labeled by bounding boxes with the goal of evaluating the performance of tracking algorithms.

- CAVIAR[2]: This dataset contains more than 80 staged indoor videos representing all kinds of human behavior such as walking, browsing, shopping, fighting, etc. Like the PETS dataset, a bounding box is associated with each moving character.

- i-LIDS[3]: This dataset contains 4 scenarios (parked vehicle, abandoned object, people walking in a restricted area, doorway). Due to the size of the videos (more than 24 hours of footage) the videos are not fully labeled.

- ETISEO[4]: This dataset contains more than 80 video clips of various indoor and outdoor scenes. Since the ground truth consists mainly of high-level information such as the bounding box, object class, event type, etc., this dataset is more suitable for tracking, classification and event recognition than change detection.

- VSSN 2006[5]: This dataset contains 9 semi-synthetic videos composed of a real background and artificially-moving objects. The videos contain animated background, illumination changes and shadows, however include no frames void of activity.

- IBM[6]: This dataset contains 15 indoor and outdoor videos taken from PETS 2001 plus additional videos. For each video, 1 frame out of 30 is labeled with a bounding box around each foreground moving object.

Further details about these datasets, and many others, can be found on a web page of the European CANTATA project[7]. With the exception of the Wallflower and VSSN 2006 datasets, all others have ground-truth information represented in terms of the bounding box for each foreground object. Furthermore, the focus in the above datasets is more on tracking as well as human behavior and interaction recognition than change detection. As such, the above datasets do not contain the diversity of video categories present in the new dataset.

[2] http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1
[3] http://www.homeoffice.gov.uk/science-research/hosdb/i-lids
[4] http://www-sop.inria.fr/orion/ETISEO
[5] http://mmc36.informatik.uni-augsburg.de/VSSN06_OSAC
[6] http://www.research.ibm.com/peoplevision/performanceevaluation.html
[7] http://www.hitech-projects.com/euprojects/cantata/datasets_cantata/

2.2. Survey Papers

Below, we list key survey papers that are devoted to the comparison and ranking of change and motion detection algorithms. Each paper uses its own dataset.

- Benezeth *et al.*, 2010 [4] use a collection of 29 videos (15 camera-captured, 10 semi-synthetic, and 4 synthetic) taken from PETS 2001, the IBM dataset, and the VSSN 2006 dataset. The authors also use semi-synthetic videos composed of synthetic foreground objects (people and cars) moving over a camera-captured background.
- Bouwmans *et al.*, 2008 [5] survey only GMM methods and use the Wallflower dataset.
- Nascimento and Marques, 2006 [16] use a single PETS 2011 video sequence which they manually label at pixel resolution using a graphical editor.
- Brutzer *et al.*, 2011 [6] use a synthetic (computer-generated) dataset produced from only one 3D scene representing a street corner. The sequences include illumination changes, dynamic background, shadows and noise, while lacking frames with no activity.
- Prati *et al.*, 2001 [19] use indoor sequences containing one moving person that are manually segmented into foreground (human), shadow, and background areas. Only 112 frames have ground-truth labels.
- Rosin and Ioannidis, 2003 [21] use a labeling program that automatically locates moving objects based on their position in space and properties such as color, size, shape, etc. These properties were not used by the change detection algorithms tested. However, the videos used are not realistic as they are limited to lab scenes with balls rolling on the floor.
- Bashir and Porikli, 2006 [3] conduct a performance evaluation of tracking algorithms using the PETS 2001 dataset by comparing the detected bounding box locations with the ground-truth.

At a high level, the existing surveys suffer from three main limitations. First, the statistics reported in these papers have not been computed on a well-balanced dataset composed of real (camera-captured) videos. Typically, synthetic videos, real videos with synthetic moving objects pasted in, or real videos out of which only 1 frame has been manually segmented for ground truth are used. Furthermore, very few datasets contain more than 10 videos. Secondly, none of the papers are accompanied by a fully-operational web site that allows users to upload their results and compare them against those of others. Thirdly, the survey papers often report common, fairly simple motion detection methods, and do not report the performance of more complex methods.

3. New Dataset: CDnet

CDnet, presented at the IEEE Change Detection Workshop [1], consists of 31 videos depicting indoor and outdoor scenes with boats, cars, trucks, and pedestrians that have been captured in different scenarios and contain a range of challenges. The videos have been obtained with different cameras ranging from low-resolution IP cameras, through mid-resolution camcorders and PTZ cameras, to thermal cameras. As a consequence, spatial resolutions of the videos in CDnet vary from 320×240 to 720×576. Also, due to diverse lighting conditions present and compression parameters used, the level of noise and compression artifacts varies from one video to another. The length of the videos also varies from 1,000 to 8,000 frames and the videos shot by low-end IP cameras suffer from noticeable radial distortion. Different cameras may have different hue bias (due to different white balancing algorithms employed) and some cameras apply automatic exposure adjustment resulting in global brightness fluctuations in time. We believe that the fact that our videos have been captured under a range of settings will help prevent this dataset from favoring a certain family of change detection methods over others.

The videos are grouped into six categories according to the type of challenge each represents. We selected videos so that the challenge in one category is unique to that category. For example, only videos in the "Shadows" category contain strong shadows and only those in the "Dynamic Background" category contain strong parasitic background motion. Such a grouping is essential for a clear identification of the strengths and weaknesses of different change detection methods. With the exception of one video in the "Baseline" category, that comes from the PETS 2006 dataset, all the videos have been captured by the authors.

3.1. Video Categories

31 videos totaling nearly 90,000 frames are grouped into the following 6 categories (Fig. 1) that have been selected to cover a wide range of change detection challenges that are representative of typical visual data captured today in surveillance, smart environment, and video analytics applications:

1. **Baseline:** This category contains four videos, two indoor and two outdoor. These videos represent a mixture of mild challenges typical of the next 4 categories. Some videos have subtle background motion, others have isolated shadows, some have an abandoned object and others have pedestrians that stop for a short while and then move away. These videos are fairly easy, but not trivial, to process, and are provided mainly as reference.

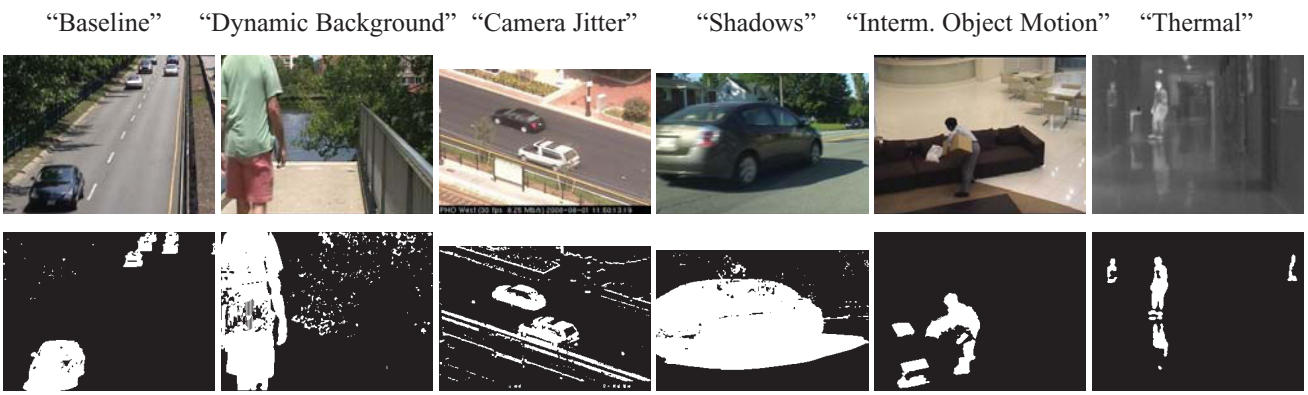

Figure 1. Sample video frames from each of the 6 categories in the new dataset available at www.changedetection.net and typical detection results obtained using basic background subtraction [4] reported in the last row of Table 1.

2. **Dynamic Background:** There are six videos in this category depicting outdoor scenes with strong (parasitic) background motion. Two videos represent boats on shimmering water, two videos show cars passing next to a fountain, and the last two depict pedestrians, cars and trucks passing in front of a tree shaken by the wind (second column in Fig. 1).

3. **Camera Jitter:** This category contains one indoor and three outdoor videos captured by unstable (e.g., vibrating) cameras. The jitter magnitude varies from one video to another.

4. **Shadows:** This category consists of two indoor and four outdoor videos exhibiting strong as well as faint shadows. Some shadows are fairly narrow while others occupy most of the scene. Also, some shadows are cast by moving objects while others are cast by trees and buildings.

5. **Intermittent Object Motion:** This category contains six videos with scenarios known for causing "ghosting" artifacts in the detected motion, i.e., objects move, then stop for a short while, after which they start moving again. Some videos include still objects that suddenly start moving, e.g., a parked vehicle driving away, and also abandoned objects. This category is intended for testing how various algorithms adapt to background changes. One example of such a challenge is shown in the 5-th column of Fig. 1 where new objects are added to or existing objects are removed from the scene.

6. **Thermal:** In this category, five videos (three outdoor and two indoor) have been captured by far-infrared cameras. These videos contain typical thermal artifacts such as heat stamps (e.g., bright spots left on a seat after a person gets up and leaves), heat reflection on floors and windows (see the last column of Fig. 1), and camouflage effects, when a moving object has the same temperature as the surrounding regions.

We would like to mention that although camouflage, caused by moving objects that have very similar color/texture to the background, is among the most glaring change detection issues, we have not created a camouflage category. This is partially because almost every real video sequence contains some level of camouflage. It is difficult to create a dataset in which there is a category exclusively with camouflage challenges while other categories are void of it.

3.2. Ground-Truth Labels

As mentioned in Section 2, the current online datasets have been designed mainly for testing tracking and scene understanding algorithms, and thus the ground truth is provided in the form of bounding boxes. Although this can be used to validate change detection methods, a precise validation requires ground truth at pixel resolution. Therefore, ideally, videos should be labeled a number of times by different persons and the results averaged out. This, however, is impractical due to resource and time constraints. Furthermore, it is very difficult for a person to produce uncontroversial binary ground-truth images for camera-captured videos. This is particularly difficult near moving object boundaries and in semi-transparent areas. Due to motion blur and partially-opaque objects (e.g., sparse bushes, dirty windows, fountains), pixels in these areas may contain both the moving object and background. As a consequence, one cannot reliably classify such pixels as belonging to either *Static* or *Moving* class. Since these areas carry a certain level of uncertainty, evaluation metrics should not be computed for pixels in these areas. Therefore, we decided to produce ground-truth images with the following labels:

Static: assigned grayscale value of 0,

Shadow: assigned grayscale value of 50,

Non-ROI[8]: assigned grayscale value of 85,

Unknown: assigned grayscale value of 170,

Moving assigned grayscale value of 255.

The *Static* and *Moving* classes are associated with pixels for which the motion status is obvious. The *Shadow* label is associated with hard and well-defined moving shadows such as the one in Fig. 2. Hard shadows are among the most difficult artifacts to cope with and we believe that adding this extra information improves the richness and utility of the dataset. Please note that evaluation metrics discussed in Section 3.3 consider the *Shadow* pixels as *Static* pixels. The *Unknown* label is assigned to pixels that are half-occluded and those corrupted by motion blur. All pixels located close to moving-object boundaries are automatically labeled as *Unknown* (Fig. 2). This prevents evaluation metrics from being corrupted by pixels whose status is unclear.

The *Non-ROI* (not in region of interest) label serves two purposes. Firstly, since most change detection methods incur a delay before their background model stabilizes, we labeled the first few hundred frames of each video sequence as *Non-ROI*. This prevents the corruption of evaluation metrics due to errors during initialization. Secondly, the *Non-ROI* label prevents the metrics from being corrupted by activities unrelated to the category considered. An example of this situation is shown in the second row of Fig. 2, which illustrates a sequence of cars that arrive, stop at a street light and then move away. The goal of the video is to measure how well a change detection method can handle intermittent motion. However, since the scene is cluttered with unrelated activities (cars on the highway) the *Non-ROI* label puts the focus on street-light activities. Similarly, the top row in Fig. 2 illustrates the *Shadow* category; the *Non-ROI* label is used to prevent the metrics from corruption by trees moving in the background.

3.3. Evaluation Metrics

Finding the right metric to accurately measure the ability of a method to detect motion or change without producing excessive false positives and false negatives is not trivial. For instance, recall favors methods with a low False Negative Rate. On the contrary, specificity favors methods with a low False Positive Rate. Having the entire precision-recall tradeoff curve or the ROC curve would be ideal, but not all methods have the flexibility to sweep through the complete gamut of tradeoffs. In addition, one cannot, in general, rank-order methods based on a curve. We deal with these difficulties by reporting the average performance of each method for each video category with respect to 7 different performance metrics each of which has been well-studied in the literature. Specifically, for each method, each video category, and each metric, we report the performance (as measured by the value of the metric) of the method averaged across all the videos of the category.

[8]ROI stands for Region of Interest.

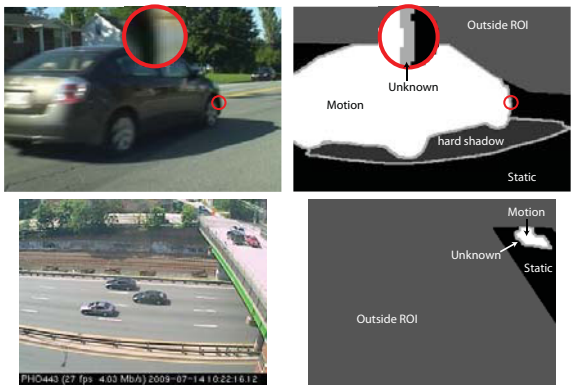

Figure 2. Sample video frames from the *Bungalows* and *Street light* sequences and corresponding 5-class ground-truth label fields.

Let TP = number of true positives, TN = number of true negatives, FN = number of false negatives, and FP = number of false positives. The 7 metrics that we use are:

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: $2\frac{Pr \cdot Re}{Pr+Re}$

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

For each method, the above metrics are first computed for each video in each category. For example, the recall metric for a particular video v in a category c is computed as follows:

$$\mathrm{Re}_{v,c} = TP_{v,c}/(TP_{v,c}+FN_{v,c}).$$

Then, a category-average metric for each category is computed from the values of the metric for all videos in a single category. For example, the average recall metric of category c is given by

$$\mathrm{Re}_c = \frac{1}{|N_c|}\sum_v \mathrm{Re}_{v,c}$$

where $|N_c|$ is the number of videos in category c. We also report an overall-average metric which is the simple average

of the category-averages. For example, the overall-average recall is given by

$$\text{Re} = \frac{1}{6}\sum_{c} \text{Re}_c. \quad (1)$$

Similar category-average and overall-average values are also computed for the other metrics and categories accordingly. The overall-average metrics such as Re are reported in Table 1 while category-average metrics such as Re_c are reported on our website. Averaging metrics in this way (as opposed to pooling together all pixels across all videos and/or categories and then averaging) prevents bias that would occur should some videos be much larger in terms of frame size and/or length; summing up across videos would give overwhelming importance to larger videos.

In order to rank-order different change detection methods, we need to rationally combine the performance across different metrics (and/or categories) into a single rank that is indicative of how well a method fares *relative* to other methods in each category and across all categories. To this end, motivated by the approach followed by Young and Ferryman [27], we provide an average ranking R across all overall-average metrics, and an average ranking RC across all categories. To explain how these are computed, let $rank_i(m, c)$ denote the rank of method i for metric m in category c. The average ranking of method i in category c across all metrics is given by:

$$\text{RM}_{c,i} = \frac{1}{7}\sum_{m} rank_i(m, c).$$

The overall ranking across categories RC_i of method i is then computed by taking the simple average of its average rankings across all 6 categories:

$$\text{RC}_i = \frac{1}{6}\sum_{c} \text{RM}_{c,i}.$$

The average ranking R_i for method i across all overall-average metrics is given by

$$\text{R}_i = \frac{1}{7}\sum_{m'} rank_i(m')$$

where m' is an overall-average metric such as the one computed in equation (1) and $rank_i(m')$ denotes the rank of method i according to the overall-average metric m'. We report the values of R, RC, and the 7 overall-average metrics for different methods in Table 1. The category-wide overall rankings and category-average metrics are available on the www.changedetection.net website.

4. Methods Tested

A total of 18 change detection methods were evaluated for the IEEE Change Detection Workshop [1]. Of these, 3 are relatively simple methods that rely on plain background subtraction, of which two use color features (the Euclidean and Mahalanobis distance methods described in [4]) and one uses local self-similarity features [11]. Two fairly old, but frequently-cited, methods: KDE-based estimation by Elgammal *et al.* [8] and GMM by Stauffer and Grimson [24], as well as 5 improved versions of these methods: self-adapting GMM by KaewTraKulPong [12], improved GMM by Zivkovic and van der Heijden [28], block-based GMM by Dora (RECTGAUSS-Tex) *et al.* [20], multi-level KDE by Nonaka *et al.* [17], and spatio-temporal KDE by Yoshinaga *et al.* [1] were also tested. We also include results for a machine learning method based on neural maps (SOBS and SC-SOBS) by Maddalena *et al.* [13, 14], a post-processing method based on probabilistic super-pixels (PSP-MRF) [22], a fairly complex commercial method that does pixel-level detection using the Chebyshev inequality and peripheral and recurrent motion detectors by Morde *et. al.* [15] and 3 stochastic methods based on background sample selection namely ViBe [2], ViBe+ [7], and Hofmann's self-adaptive method (PBAS) [10]. We also included a per-pixel recursive Bayesian background method, which shows the best robustness to shadows [18], in our evaluations.

Out of the above methods, all except for the Euclidean and Mahalanobis distance methods [4], are robust to background motion, four are robust to shadows [13, 28, 12, 18] and two are robust to artifacts stemming from intermittent motion [2, 7].

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 our website.

5. Experimental Results

The overall results are shown in Table 1 where the methods have been sorted according to their average ranking across categories (RC). A more comprehensive tabulation of performance can be found on our website, where a visitor can re-sort the methods by the average overall ranking R as well as individual average metrics.

It should come as no surprise that the three simplest methods based on plain background subtraction [4, 11] are at the bottom of the table, whereas the four most recent methods [22, 7, 10, 14] are at the top. The methods ranked number 1 [10] and number 3 [7] are closely related. Both methods make use of a non-parametric probabilistic model for the background at each spatial location based on a random subset of pixel values from the recent past. Such a stochastic non-parametric model makes these methods robust to instabilities (background motion and camera jitter) and intermittent motion artifacts. The success of the number 1 [10] method can be attributed to the use of a dynamic

Method	RC	R	Re	Sp	FPR	FNR	PWC	F-Measure	Pr
PBAS [10]	3.00	3.29	0.78	0.990	0.010	0.009	1.77	0.75	0.82
PSP-MRF [22]	4.83	5.71	0.80	0.983	0.017	0.009	2.39	0.74	0.75
ViBe+ [7]	4.83	5.00	0.69	0.993	0.007	0.017	2.18	0.72	0.83
SC-SOBS [14]	6.00	6.14	0.80	0.983	0.017	0.009	2.41	0.73	0.73
Chebyshev probability [15]	6.67	5.86	0.71	0.989	0.011	0.015	2.39	0.70	0.79
SOBS [13]	8.17	8.57	0.79	0.982	0.018	0.009	2.56	0.72	0.72
KDE Nonaka *et al.* [17]	9.17	8.43	0.65	0.993	0.007	0.025	2.89	0.64	0.77
ViBe [2]	9.33	10.71	0.68	0.983	0.017	0.018	3.12	0.67	0.74
GMM KaewTraKulPong [12]	9.50	9.43	0.51	0.995	0.005	0.029	3.11	0.59	0.82
KDE Elgammal [8]	9.67	11.43	0.74	0.976	0.024	0.014	3.46	0.67	0.68
KDE Yoshinaga *et al.* [1]	10.67	9.29	0.66	0.991	0.009	0.024	3.00	0.64	0.73
Bayesian Back [18]	11.00	12.57	0.60	0.983	0.017	0.020	3.39	0.63	0.74
GMM Stauffer-Grimson [24]	11.50	10.14	0.71	0.986	0.014	0.020	3.10	0.66	0.70
GMM Zivkovic [28]	13.67	10.86	0.70	0.984	0.016	0.019	3.15	0.66	0.71
GMM RECTGAUSS-Tex [20]	13.67	13.00	0.52	0.986	0.014	0.027	3.68	0.52	0.72
Local-Self similarity [11]	14.67	13.14	0.94	0.851	0.149	0.002	14.30	0.50	0.41
Mahalanobis distance [4]	15.50	13.43	0.76	0.960	0.040	0.011	4.66	0.63	0.60
Euclidean distance [4]	16.67	14.00	0.70	0.969	0.031	0.017	4.35	0.61	0.62

Table 1. Overall results across all categories (RC: average ranking across categories, R: average overall ranking).

control algorithm for automatically adapting thresholds and other parameter values. The second ranked method [22] is a surprisingly simple super-pixel-based post-processing method that can be combined with almost any other change detection method to improve its performance. As shown in the paper, this approach reduces both the FNR and the FPR of any method it is used on. As for the fourth ranked method SC-SOBS [14], its approach is orthogonal to traditional motion detection methods in its use of a self-organizing neural network. Such an approach gives remarkable results on baseline and intermittent object motion videos.

A bit surprising is the fact that with the exception of the method by KaewTraKulPong [12], the KDE-based methods outperform the GMM-based methods. Another interesting observation is that the F-measure correlates better with the average rankings than any of the other measures; recall, FPR, FNR and precision are much less consistent.

In order to assess the challenge that each video category poses for the tested methods, we ranked the categories according to the median metrics obtained by all methods in a given category. As can be seen in Table 2, videos with intermittent motion, shadows and camera jitter pose a greater challenge than videos in the other categories. For example, videos with intermittent motion pose the largest challenge in terms of the F-Measure (0.5), FPR (0.29) and PWC (6%). On the other hand, videos exhibiting steady background motion seem to be less challenging. Many methods had difficulty with thermal videos as most of the time they suffered from camouflage problems, resulting in large FNR scores.

It is clear from Table 2 that video sequences with strong shadows pose a significant challenge for the tested methods. In order to verify this observation, we have also computed FPR within shadow areas (FPR-S) for all videos in the "Shadows" category. As can be seen in Table 3, the tested methods attained FPR in shadow areas between 0.33 and 0.64. This large FPR indicates that most of the 18 methods cannot effectively deal with shadows.

6. Future Work

The CDnet undertaking aims to provide the research community with a rigorous and comprehensive scientific benchmarking facility, a rich dataset of videos, a set of utilities, and an access to author-approved algorithm implementations for testing and ranking of existing and new algorithms for motion and change detection.

The already extensive dataset will be regularly revised and expanded with feedback from the academia and industry. We will maintain and update a rank list of the most accurate motion and change detection algorithms in the various categories for years to come.

Acknowledgement

Janusz Konrad was partially supported for this work by the NSF Foundation under grant ECS-0905541 and Pierre-Marc Jodoin by NSERC Discovery Grant 371951.

References

[1] 1st IEEE Change Detection Workshop, 2012, in conjunction with CVPR. www.changedetection.net. 2, 3, 6, 7, 8

Category	F-Measure	FPR	FNR	PWC
Interm. Motion	0.50	0.029	0.04	6.0
Camera Jitter	0.60	0.025	0.01	4.0
Dynamic Back.	0.63	0.010	0.002	1.3
Thermal	0.66	0.004	0.03	3.2
Shadows	0.76	0.011	0.09	2.2
Baseline	0.87	0.003	0.007	1.0

Table 2. Median F-Measure, FPR, FNR and PWC obtained by all 18 methods for each category.

Method	FPR-S
Bayesian Back [18]	0.33
KDE Nonaka *et al.* [17]	0.39
KDE Yoshinaga *et al.* [1]	0.40
GMM KaewTraKulPong [12]	0.41
Chebyshev probability [15]	0.42
RECTGAUSS-Tex [20]	0.48
ViBe+ [7]	0.53
GMM Stauffer-Grimson [24]	0.54
GMM Zivkovic [28]	0.54
ViBe [2]	0.55
SOBS [13]	0.57
Euclidean distance [4]	0.58
PBAS [10]	0.58
PSP-MRF [22]	0.59
Mahalanobis distance [4]	0.59
SC-SOBS [14]	0.60
KDE Elgammal [8]	0.62
Local-Self similarity [11]	0.64

Table 3. Ranking of methods according to FPR in shadow areas for videos in "Shadows" category.

[2] O. Barnich and M. Van Droogenbroeck. ViBe: A universal background subtraction algorithm for video sequences. *IEEE Trans. Image Process.*, 20(6):1709–1724, 2011. 6, 7, 8

[3] F. Bashir and F. Porikli. Performance evaluation of object detection and tracking systems. In *Proc. IEEE Int. Workshop on Vis. Surv. and Perf. Eval. of Tracking and Surv.*, 2006. 3

[4] 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. 3, 4, 6, 7, 8

[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] S. Brutzer, B. Hferlin, and G. Heidemann. Evaluation of background subtraction techniques for video surveillance. In *Proc. IEEE Conf. Computer Vision Pattern Recognition*, pages 1937–1944, 2011. 3

[7] M. V. Droogenbroeck and O. Paquot. Background subtraction: Experiments and improvements for ViBe. In *IEEE Workshop on Change Detection*, 2012. 6, 7, 8

[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. 6, 7, 8

[9] A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. *Proc. European Conf. on Computer Vision,* Dublin, Ireland, 2000. 1

[10] M. Hofmann. Background segmentation with feedback: The pixel-based adaptive segmenter. In *IEEE Workshop on Change Detection*, 2012. 6, 7, 8

[11] J.-P. Jodoin, G. Bilodeau, and N. Saunier. Background subtraction based on local shape. Technical Report arXiv:1204.6326v1, 2012. 6, 7, 8

[12] P. KaewTraKulPong and R. Bowden. An improved adaptive background mixture model for realtime tracking with shadow detection. *European Workshop on Advanced Video Based Surveillance Systems*, 2001. 6, 7, 8

[13] L. Maddalena and A. Petrosino. A self-organizing approach to background subtraction for visual surveillance applications. *IEEE Transactions on Image Processing*, 17(7):1168–1177, 2008. 6, 7, 8

[14] L. Maddalena and A. Petrosino. The SOBS algorithm: what are the limits? In *IEEE Workshop on Change Detection*, 2012. 6, 7, 8

[15] A. Morde, X. Ma, and S. Guler. Learning a background model for change detection. In *IEEE Workshop on Change Detection*, 2012. 6, 7, 8

[16] J. Nascimento and J. Marques. Performance evaluation of object detection algorithms for video surveillance. *IEEE Trans. Multimedia*, 8(8):761–774, 2006. 3

[17] Y. Nonaka, A. Shimada, H. Nagahara, and R. Taniguchi. Evaluation report of integrated background modeling based on spatio-temporal features. In *IEEE Workshop on Change Detection*, 2012. 6, 7, 8

[18] F. Porikli and O. Tuzel. Bayesian background modeling for foreground detection. *Proc. of ACM Visual Surveillance and Sensor Network*, 2005. 6, 7, 8

[19] 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 Pattern Recognition*, pages 571–577, 2001. 3

[20] D. Riahi, P. St-Onge, and G. Bilodeau. RECTGAUSS-tex: Block-based background subtraction. Technical Report EPM-RT-2012-03, Ecole Polytechnique de Montreal, 2012. 6, 7, 8

[21] P. Rosin and E. Ioannidis. Evaluation of global image thresholding for change detection. *Pattern Recognit. Lett.*, 24:2345–2356, 2003. 3

[22] A. Schick, M. Bäuml, and R. Stiefelhagen. Improving foreground segmentations with probabilistic superpixel markov random fields. In *IEEE Workshop on Change Detection*, 2012. 6, 7, 8

[23] C. Stauffer and E. Grimson. Adaptive background mixture models for real-time tracking. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition,* Fort Collins, CO, 1999. 1

[24] C. Stauffer and E. Grimson. Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(8):747–757, 2000. 6, 7, 8

[25] 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

[26] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. Pattern Anal. Machine Intell.*, 19(7):780–785, 1997. 1

[27] D. Young and J. Ferryman. PETS metrics: Online performance evaluation service. In *Proc. IEEE Int. Workshop on Vis. Surv. and Perf. Eval. of Tracking and Surv.*, pages 317–324, 2005. 2, 6

[28] Z. Zivkovic and F. V. D. Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognit. Lett.*, 27:773–780, 2006. 6, 7, 8