US010706558B2

(12) **United States Patent**
Van Droogenbroeck et al.

(10) **Patent No.:** **US 10,706,558 B2**
(45) **Date of Patent:** **Jul. 7, 2020**

(54) **FOREGROUND AND BACKGROUND DETECTION METHOD**

(71) Applicant: **Université de Liège**, Angleur (BE)

(72) Inventors: **Marc Van Droogenbroeck**, Overijse (BE); **Marc Braham**, Herve (BE); **Anthony Cioppa**, Oupeye (BE)

(73) Assignee: **UNIVERSITÉ DE LIÈGE**, Angleur (BE)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **16/288,468**

(22) Filed: **Feb. 28, 2019**

(65) **Prior Publication Data**

US 2019/0197696 A1 Jun. 27, 2019

Related U.S. Application Data

(63) Continuation-in-part of application No. 16/267,474, filed on Feb. 5, 2019, which is a continuation-in-part of application No. 15/983,238, filed on May 18, 2018, now Pat. No. 10,614,736.

(30) **Foreign Application Priority Data**

Aug. 4, 2017 (EP) 17184931

(51) **Int. Cl.**
G06T 7/194	(2017.01)
G06T 7/215	(2017.01)
G06T 7/70	(2017.01)
G06T 7/254	(2017.01)
(Continued)	

(52) **U.S. Cl.**
CPC *G06T 7/194* (2017.01); *G06K 9/726* (2013.01); *G06T 7/11* (2017.01); *G06T 7/215*

(2017.01); *G06T 7/254* (2017.01); *G06T 7/70* (2017.01); *G06T 2207/10016* (2013.01);
(Continued)

(58) **Field of Classification Search**
CPC . G06T 7/194; G06T 7/11; G06T 7/254; G06T 7/70; G06T 7/215; G06T 7/251; G06T 2207/10024; G06T 2207/20224; G06T 2207/20076; G06T 2207/10016;
(Continued)

(56) **References Cited**

U.S. PATENT DOCUMENTS

9,042,662	B2 *	5/2015	Aerts	G06T 7/11
				382/224
2006/0045354	A1 *	3/2006	Hanna	G06K 9/00771
				382/224

(Continued)

OTHER PUBLICATIONS

P-M. Jodoin et al., "Overview and Benchmarking of Motion Detection Methods", Background Modeling and Foreground Detection for Video Surveillance, Chapter 24, pp. 1-1-1-26, Jul. 2014, cited in the Specification.

(Continued)

Primary Examiner — John B Strege
(74) *Attorney, Agent, or Firm* — Jacobson Holman, PLLC.

(57) **ABSTRACT**

The present invention concerns a method for assigning a pixel to one of a foreground pixel set and a background pixel set using asynchronous semantic post processing to improve motion detection by an imaging device, the pixel belonging to an image of a chronological sequence of images taken by the imaging device that includes background and foreground objects.

16 Claims, 12 Drawing Sheets
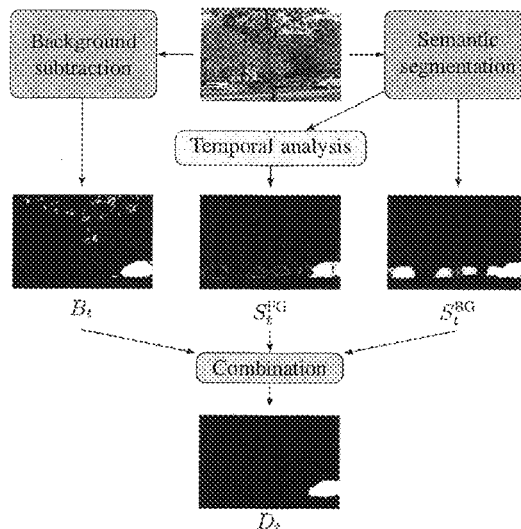
(51) **Int. Cl.**
    G06K 9/72   (2006.01)
    G06T 7/11   (2017.01)

(52) **U.S. Cl.**
    CPC *G06T 2207/10024* (2013.01); *G06T 2207/20076* (2013.01); *G06T 2207/20224* (2013.01)

(58) **Field of Classification Search**
    CPC G06T 2207/20084; G06K 9/726; G06K 9/6277
    See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2011/0088019 A1*	4/2011	Jedlicka	G06F 8/33 717/143
2012/0327172 A1*	12/2012	El-Saban	G06K 9/00228 348/14.02
2015/0146917 A1*	5/2015	Bernal	G06K 9/00771 382/103
2015/0187076 A1*	7/2015	Lin	H04N 7/15 382/103
2017/0300830 A1*	10/2017	Kadav	G06F 15/17312
2018/0211393 A1*	7/2018	Chen	G06T 7/194

OTHER PUBLICATIONS

T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: An overview", Computer Science Review, vol. 11-12, pp. 31-66, May 2014, cited in the Specification.

C. Stauffer et al., "Adaptive background mixture models for real-time tracking", IEEE Int. Conf. Comput. Vision and Pattern Recogn. (CVPR), vol. 2, pp. 246-252, Jun. 1999, cited in the Specification.

O. Barnich et al., "ViBe: A universal background subtraction algorithm for video sequences", IEEE Trans. Image Process., vol. 20, No. 6, pp. 1709-1724, Jun. 2011, cited in the Specification.

V. Jain et al., "Background Modeling Based on Subpixel Edges", IEEE Int. Conf. Image Process. (ICIP), vol. 6, pp. 321-324, Sep. 2007, cited in the Specification.

S. Zhang et al., "Dynamic Background Modeling and Subtraction Using Spatio-Temporal Local Binary Patterns", IEEE Int. Conf. Image Process. (ICIP), pp. 1556-1559, Oct. 2008, cited in the Specification.

M. Chen et al., "Spatiotemporal Background Subtraction Using Minimum Spanning Tree and Optical Flow*", Eur. Conf. Comput. Vision (ECCV), vol. 8695 of Lecture Notes Comp. Sci., pp. 521-534, Sep. 2014, cited in the Specification.

M. Braham et al., "A Physically Motivated Pixel-Based Model for Background Subtraction in 3D Images," in IEEE Int. Conf. 3D Imaging (IC3D), pp. 1-8, Dec. 2014, cited in the Specification.

T. Bouwmans et al., "On the Role and the Importance of Features for Background Modeling and Foreground Detection", CoRR, vol. abs/1611.09099, pp. 1-131, Nov. 2016, cited in the Specification.

S. Piérard et al., "A perfect estimation of a background image does not lead to a perfect background subtraction: analysis of the upper bound on the performance", Int. Conf. Image Anal. and Process. (ICIAP), Workshop Scene Background Modeling and Initialization (SBMI). vol. 9281 of Lecture Notes Comp. Sci., pp. 527-534, Sep. 2015, cited in the Specification.

J. Long et al., "Fully Convolutional Networks for Semantic Segmentation", IEEE Int. Conf. Comput. Vision and Pattern Recogn. (CVPR), pp. 3431-3440, Jun. 2015, cited in the Specification.

S. Zheng et al., "Conditional Random Fields as Recurrent Neural Networks", IEEE Int. Conf. Comput. Vision (ICCV), pp. 1529-1537, Dec. 2015, cited in the Specification.

H. Zhao et al., "Pyramid Scene Parsing Network", CoRR, vol. abs/1612.01105, Dec. 2016, cited in the Specification.

L. Sevilla-Lara et al., "Optical Flow with Semantic Segmentation and Localized Layers", IEEE Int. Conf. Comput. Vision and Pattern Recogn. (CVPR), pp. 3889-3898, Jun. 2016, cited in the Specification.
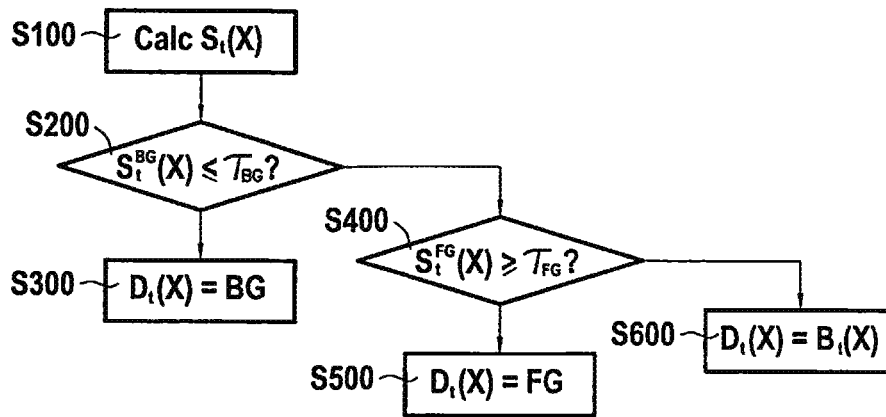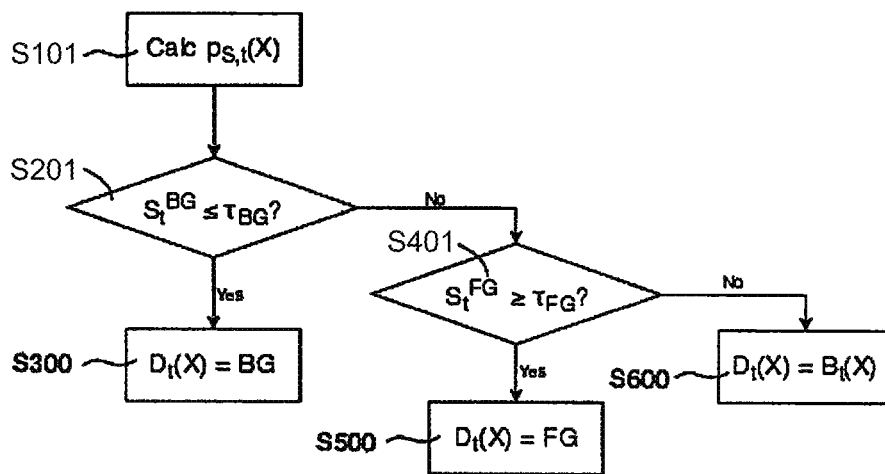
* cited by examiner

S100 — Calc $S_t(X)$

S200 — $S_t^{BG}(X) \leq \mathcal{T}_{BG}$?

S300 — $D_t(X) = BG$

S400 — $S_t^{FG}(X) \geq \mathcal{T}_{FG}$?

S500 — $D_t(X) = FG$

S600 — $D_t(X) = B_t(X)$

FIG. 1

S101 — Calc $p_{S,t}(X)$

S201 — $S_t^{BG} \leq \tau_{BG}$?

No

S401 — $S_t^{FG} \geq \tau_{FG}$?

No

Yes

S300 — $D_t(X) = BG$

Yes

S500 — $D_t(X) = FG$

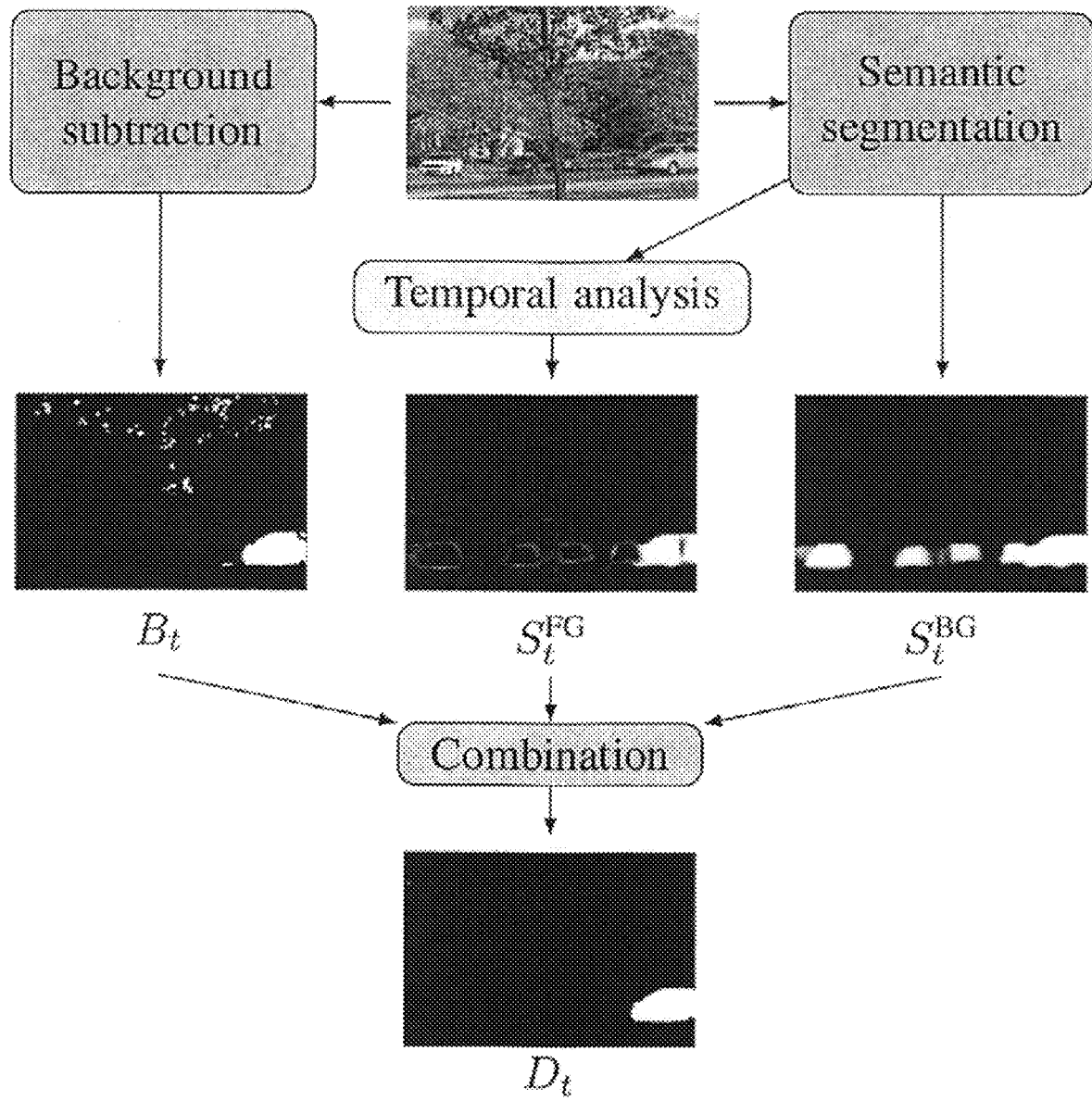S600 — $D_t(X) = B_t(X)$

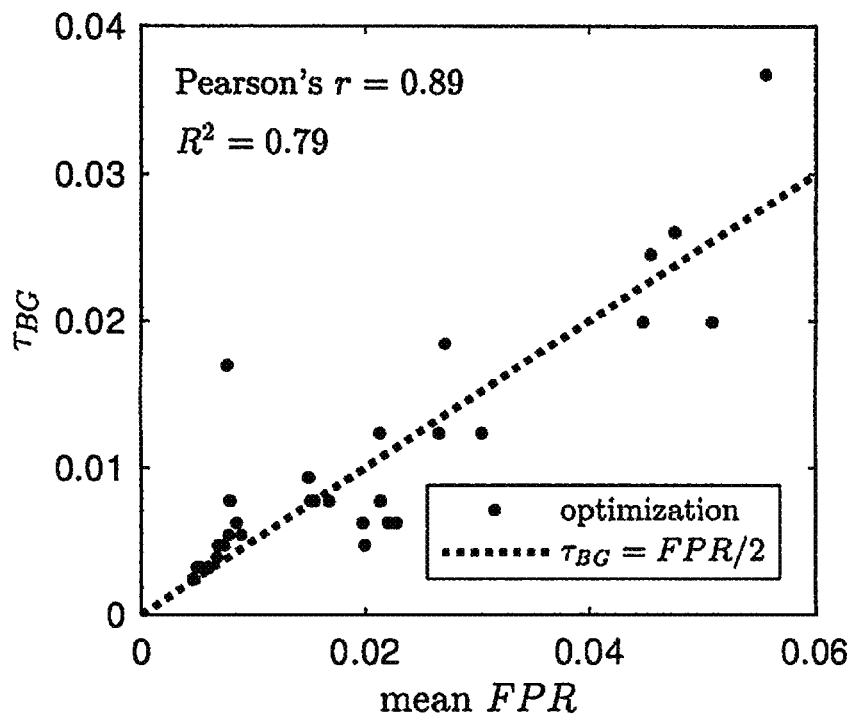## FIG. 1A

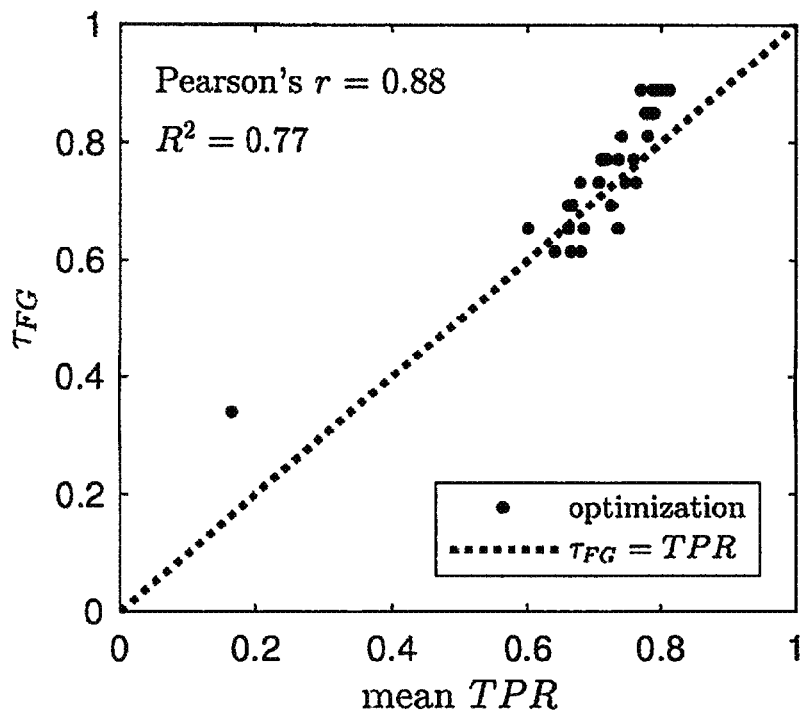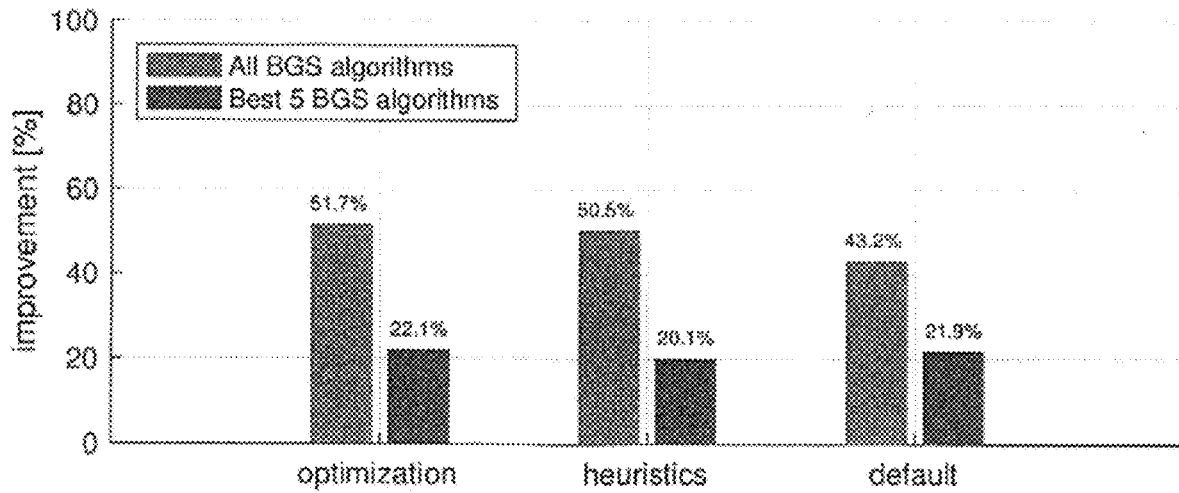B_t S_t^{FG} S_t^{BG}

D_t

FIG. 2

FIG. 3A

FIG. 3B
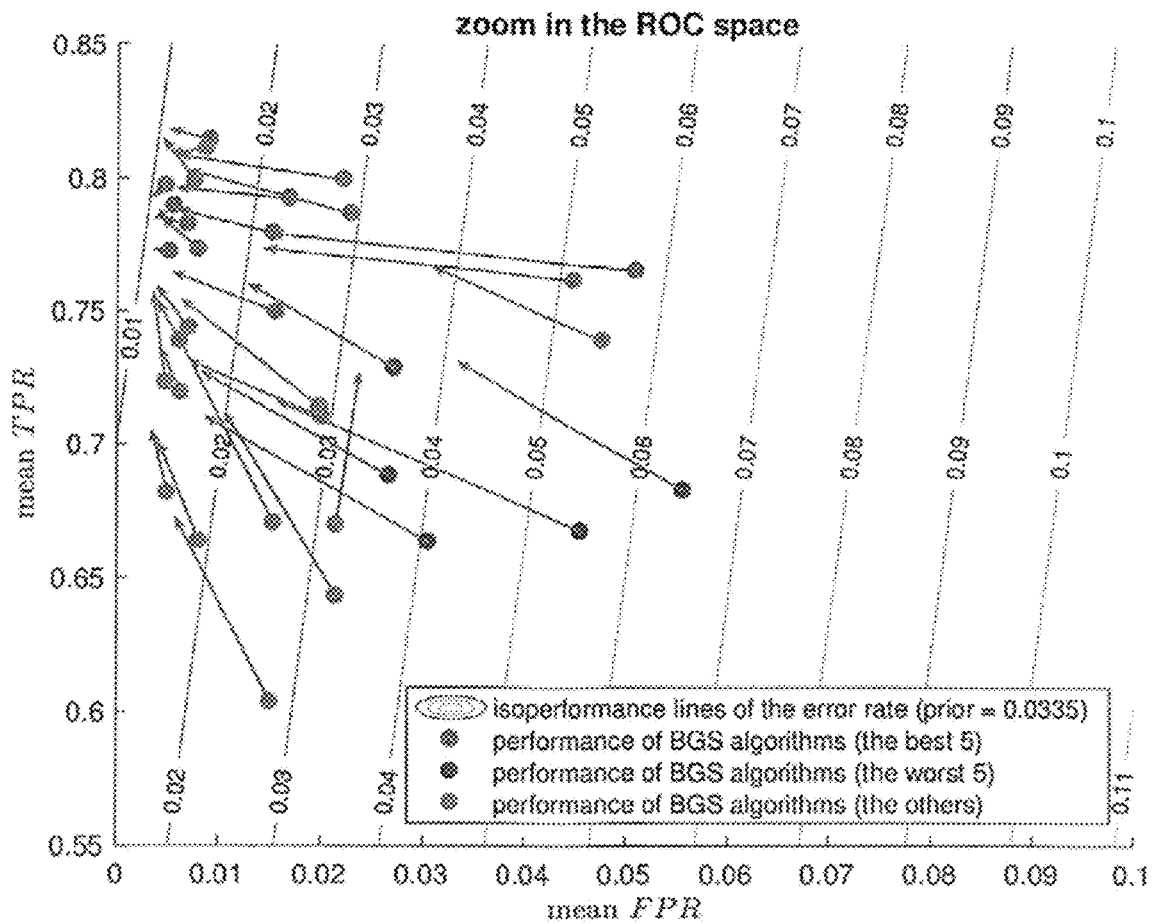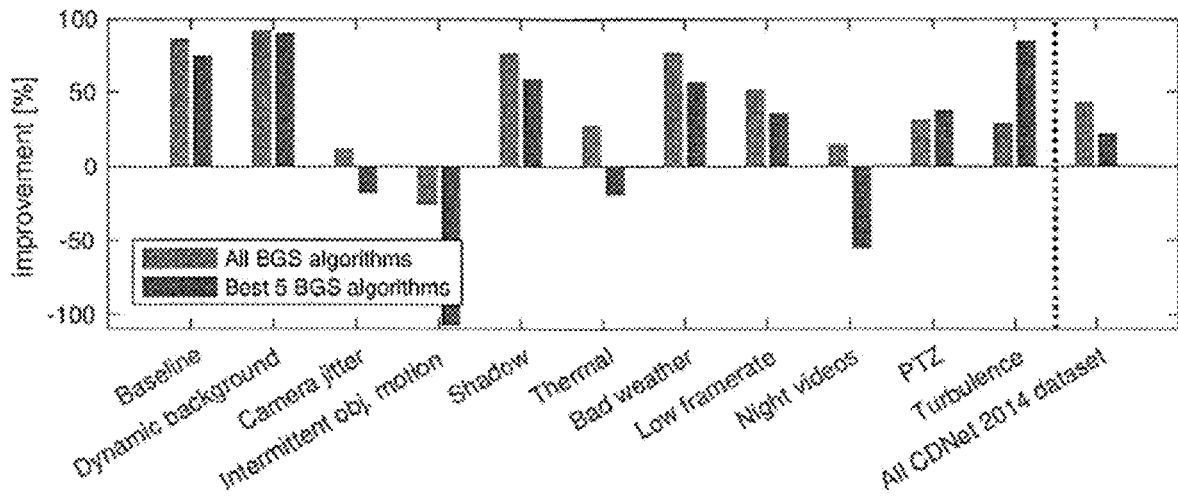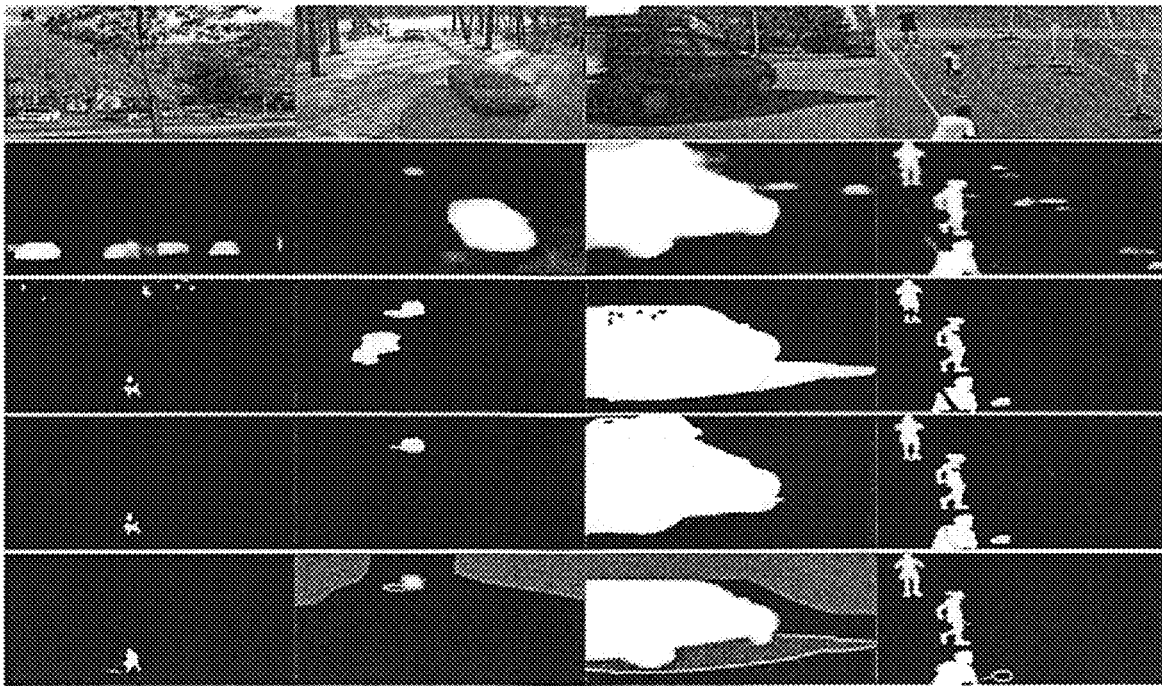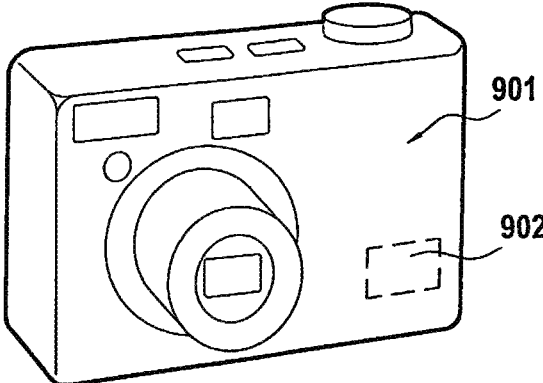
FIG. 4

FIG. 5

FIG. 6

FIG. 7

FIG. 8

FIG. 9

FIG. 10

FIG. 11

IUTIS-5

PAWCS

SuBSENSE

WeSamBe

Temporal sampling factor X:1

Temporal sampling factor X:1

Original BGS   SBS (with copy of B_t)   SBS (with copy of S_t)   ASPP without feedback

FIG. 12

Zoom in the mean ROC space

Zoom in the mean ROC space
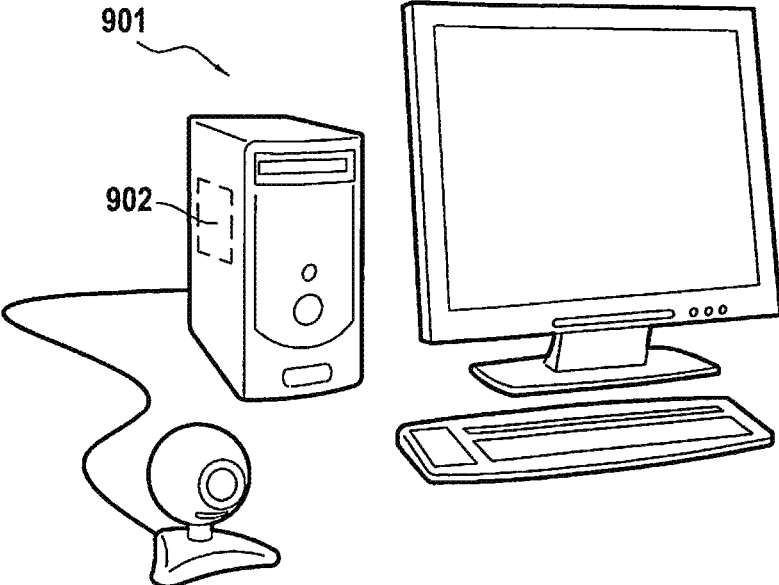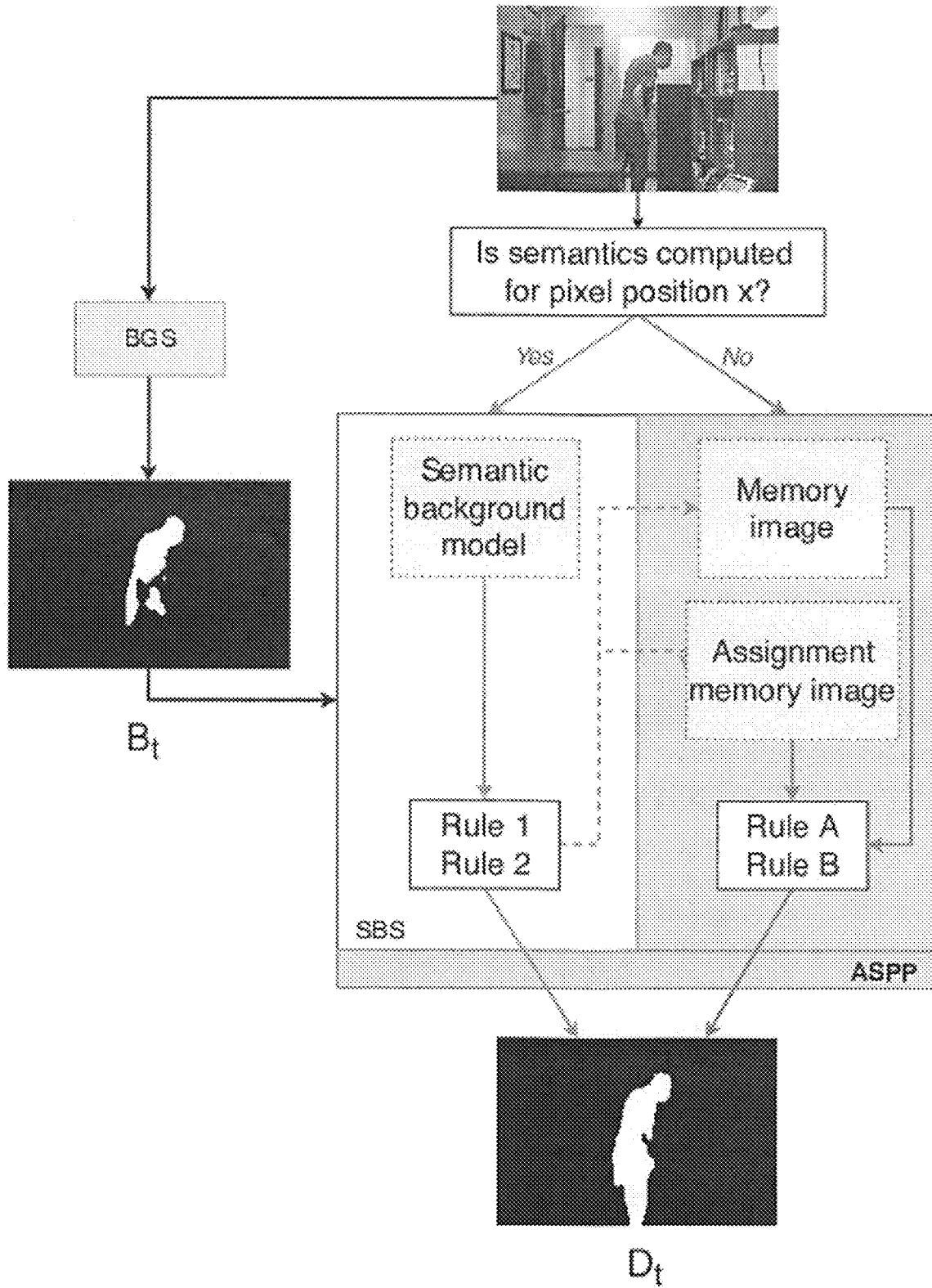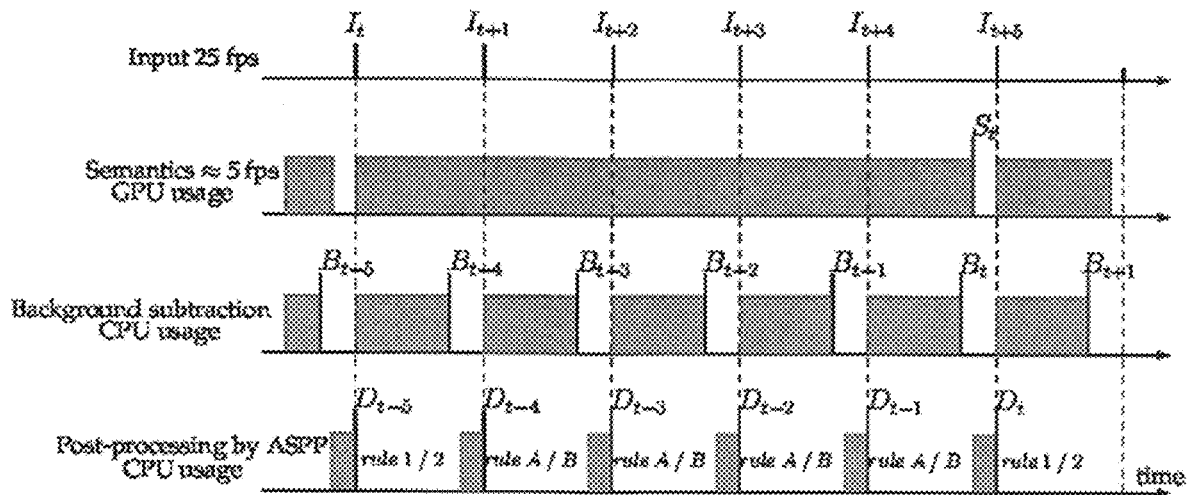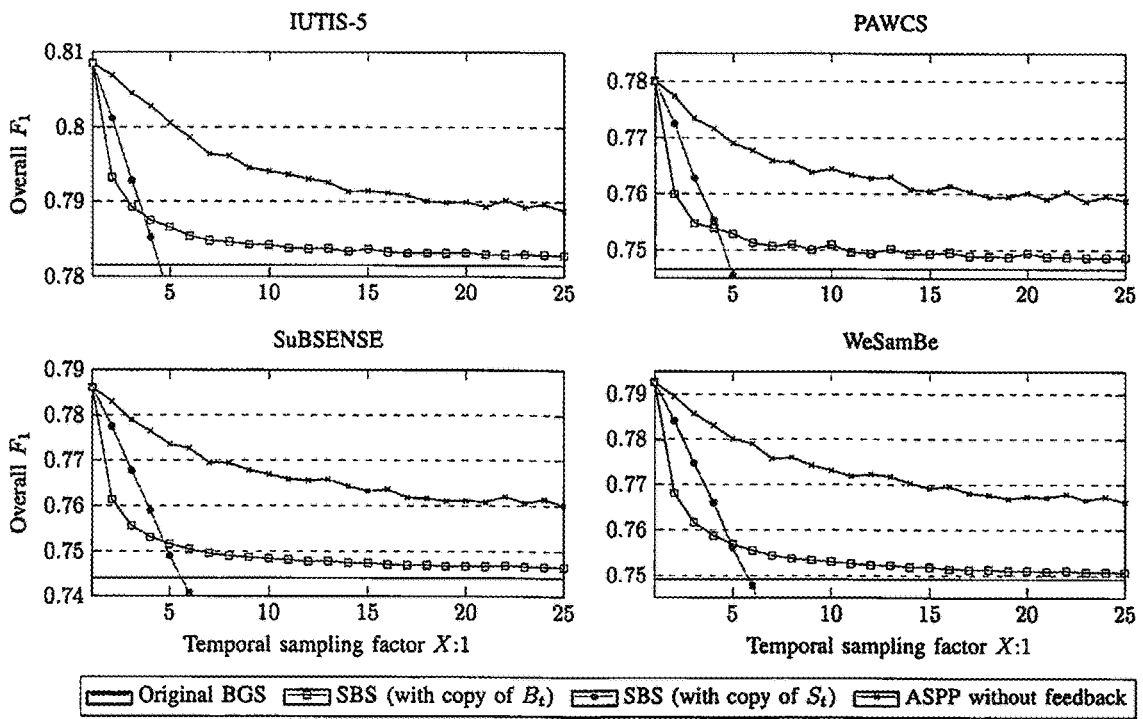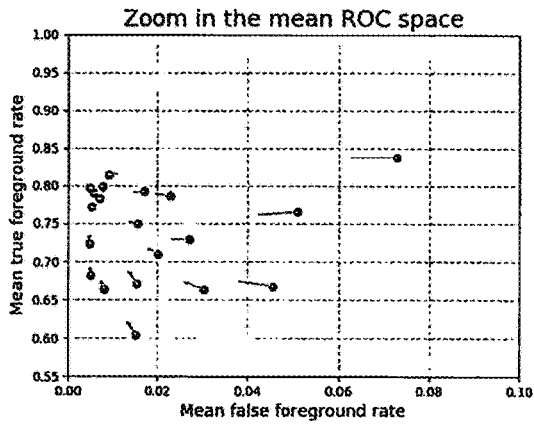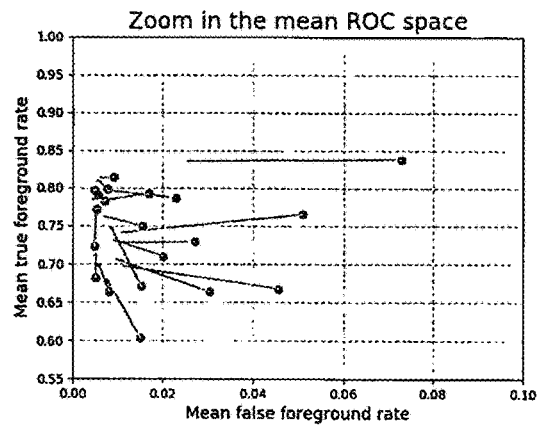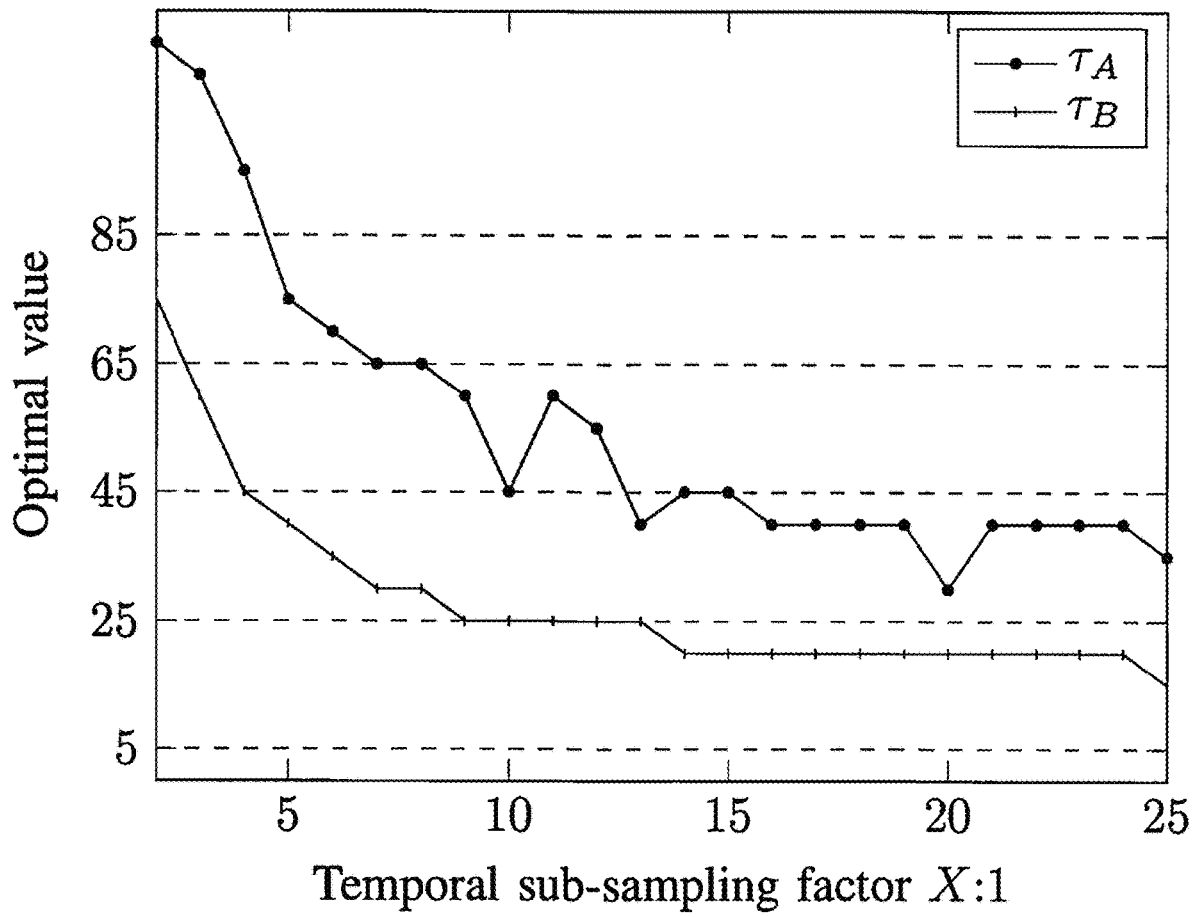
(a) SBS at 5:1

(b) ASPP at 5:1 (ours)

FIG. 13

FIG. 14

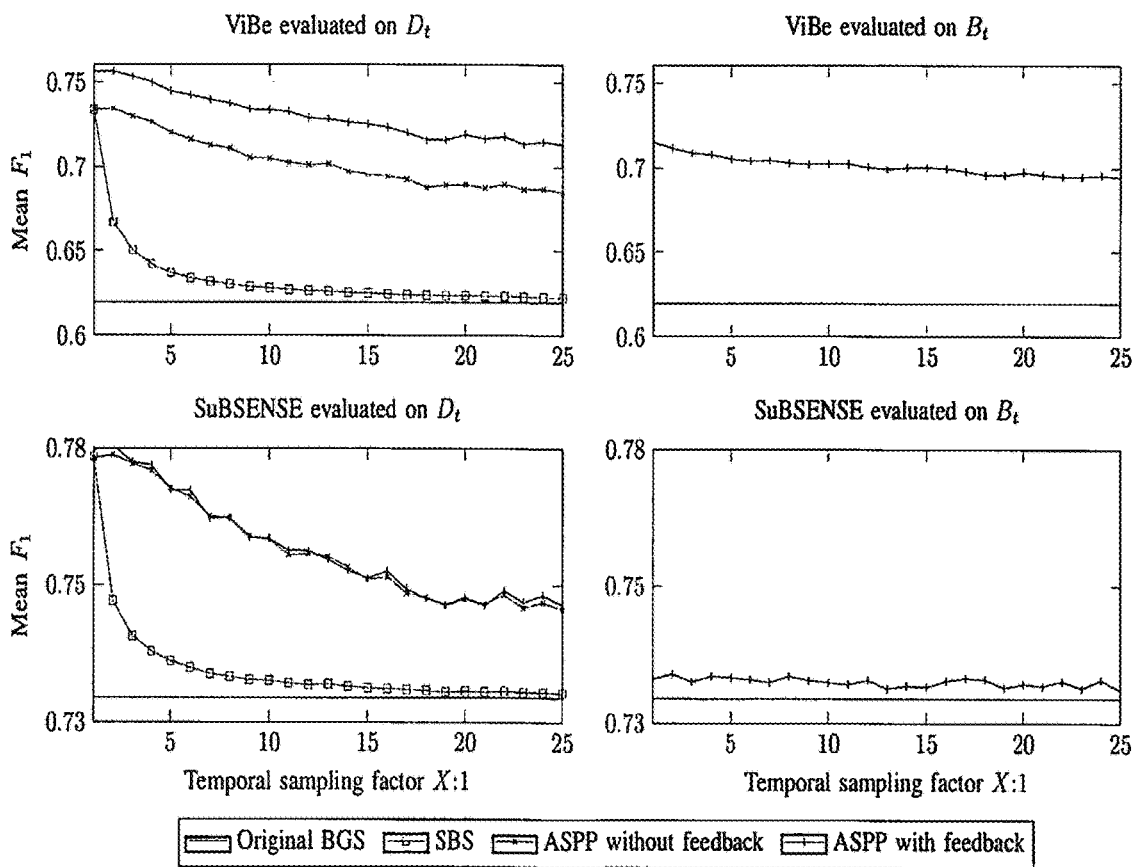FIG. 15

FOREGROUND AND BACKGROUND DETECTION METHOD

TECHNICAL FIELD

The disclosure relates to a method for assigning a pixel to one of a foreground and a background pixel sets.

BACKGROUND

A major research area in computer vision is the field of motion detection. The aim of motion detection is to classify pixels according to whether they belong to such a moving object or not, filtering any pixels that may be misclassified, so as to detect moving objects in a scene. This task, which is solved in nature with apparent ease by even rudimentary animal vision systems, has turned out to be complex to replicate in computer vision.

In the field of computer vision, an image may be expressed as a plurality of picture elements, or pixels. Each single pixel in an image may have a position x in the image and a pixel value $\vec{T}(x)$.

The position x may have any number of dimensions. For this reason, although the term "voxel" (for "volume element") is sometimes used instead of "pixel" in the field of 3D imaging, the term "pixel" should be understood broadly in the present disclosure as also covering such voxels and any picture element in images having any number of dimensions, including 3D images and/or multispectral images.

This position x may be limited to a finite domain, for instance if it is an image captured by a fixed imaging device. However, it may alternatively not be limited to a finite domain, for example if the image is captured by a moving imaging device, such as, for example, a satellite on-board camera.

The pixel value $\vec{T}(x)$ may also have any number of dimensions. For example, in a monochromatic image, the pixel value $\vec{T}(x)$ may be a scalar luminance value, but in polychromatic images, such as red-green-blue (RGB) component video images or hue saturation value (HSV) images, this pixel value $\vec{T}(x)$ may be a multidimensional vector value.

Over the last two decades, a large number of background subtraction algorithms have been proposed for motion detection. Many of these background subtraction algorithms have been reviewed by P.-M. Jodoin, S. Piérard, Y. Wang, and M. Van Droogenbroeck in "Overview and benchmarking of motion detection methods", Background Modeling and Foreground Detection for Video Surveillance, chapter 24, Chapman and Hall/CRC, July 2014, and by T. Bouwmans in "Traditional and recent approaches in background modeling for foreground detection: An overview", Computer Science Review, vol. 11-12, pp. 31-66, May 2014.

Most background subtraction algorithms involve a comparison of low-level features, such as individual pixel values, in each image, with a background model, which may be reduced to an image free of moving objects and possibly adaptive. Pixels with a noticeable difference with respect to the background model may be assumed to belong to moving objects, and may thus be assigned to a set of foreground pixels, while the remainder may be assigned to a set of background pixels. For instance, the background subtraction algorithms disclosed by C. Stauffer and E. Grimson in "Adaptive background mixture models for real-time tracking", IEEE Int. Conf. Comput. Vision and Pattern Recogn.

(CVPR), June 1999, vol. 2, pp. 246-252, and by O. Barnich and M. Van Droogenbroeck in "ViBe: A universal background subtraction algorithm for video sequences" in IEEE Trans. Image Process, vol. 20, no. 6, pp. 1709-1724, June 2011, classify pixels according to color components, whereas the background subtraction algorithms disclosed by V. Jain, B. Kimia, and J. Mundy in "Background modeling based on subpixel edges," IEEE Int. Conf. Image Process. (ICIP), September 2007, vol. 6, pp. 321-324, S. Zhang, H. Yao, and S. Liu in "Dynamic background modeling and subtraction using spatio-temporal local binary patterns", IEEE Int. Conf. Image Process. (ICIP), October 2008, pp. 1556-1559, M. Chen, Q. Yang, Q. Li, G. Wang, and M.-H. Yang in "Spatiotemporal background subtraction using minimum spanning tree and optical flow", Eur. Conf. Comput. Vision (ECCV), September 2014, vol. 8695 of Lecture Notes Comp. Sci., pp. 521-534, Springer, and M. Braham, A. Lejeune, and M. Van Droogenbroeck, "A physically motivated pixel-based model for background subtraction in 3D images," in IEEE Int. Conf. 3D Imaging (IC3D), December 2014, pp. 1-8, use, respectively, edges, texture descriptors, optical flow, or depth to assign pixels to the foreground or the background. A comprehensive review and classification of features used for background modeling was given by T. Bouwmans, C. Silva, C. Marghes, M. Zitouni, H. Bhaskar, and C. Frelicot in "On the role and the importance of features for background modeling and foreground detection," CoRR, vol. abs/1611.09099, pp. 1-131, November 2016.

While most of these low-level features can be computed with a very low computational load, they cannot address simultaneously the numerous challenges arising in real-world video sequences such as illumination changes, camouflage, camera jitter, dynamic backgrounds, shadows, etc. Upper bounds on the performance of pixel-based methods based exclusively on RGB color components were simulated by S. Piérard and M. Van Droogenbroeck in "A perfect estimation of a background image does not lead to a perfect background subtraction: analysis of the upper bound on the performance," in Int. Conf. Image Anal. and Process. (ICIAP), Workshop Scene Background Modeling and Initialization (SBMI). September 2015, vol. 9281 of Lecture Notes Comp. Sci., pp. 527-534, Springer. In particular, it was shown that such background subtraction algorithms fail to provide a perfect segmentation in the presence of noise and shadows, even when a perfect background image is available.

Among the typical challenges for background subtraction algorithms, we can in particular consider camouflaged foreground objects, "ghosts", dynamic backgrounds and shadows and/or reflection effects.

A foreground object is considered to be "camouflaged" when its corresponding pixel values (e.g. color or luminance) are similar to those of the background. In this situation, background subtraction algorithms may erroneously assign the corresponding foreground pixels to the background, as false negatives. This may for instance take the form of color camouflage on images from color cameras, or of thermal camouflage on images from thermal cameras. Snow cover, for example, may lead to such camouflaging.

"Ghosting" is the phenomenon when a previously static object, which thus belonged to the background, starts moving. In this situation, because not only the pixel values of the pixels corresponding to the object change, but also those belonging to the background previously hidden by the object when it was static, these latter background pixels may be erroneously assigned to the foreground, as false positives.

Dynamic backgrounds are such backgrounds were there may be changes in pixel values, such as for instance a windblown leafy tree or a sea wave. In this situation, the corresponding background pixels may be erroneously assigned to the foreground, also as false positives.

Similarly, shadows and reflections may lead to background pixels being erroneously assigned to the foreground, as false positives, due to the associated changes in pixel values.

Other challenges that may lead background pixels to be erroneously assigned to the foreground as false positives are noisy images (for instance due to compression artifacts), camera jitter, automatic camera adjustments, low frame rates, panning, tilting and/or zooming, bad weather, gradual or sudden lighting changes, motion/insertion of background objects, residual heat stamps on thermal images, persistent background changes, clouds, smoke and highlights due to reflections.

Other challenges that may lead foreground pixels to be erroneously assigned to the background are fast moving objects, and foreground objects that become motionless and may thus be erroneously incorporated into the background.

SUMMARY

A first aspect of the disclosure relates to a method for assigning a pixel to one of a foreground pixel set and a background pixel set, more reliably and robustly than with background subtraction algorithms comparing a pixel value of the pixel with a pixel value of a corresponding pixel in a background model. In particular, according to this first aspect, the present disclosure seeks to address the above-mentioned challenges to background subtraction algorithms. For this purpose, the method according to this first aspect may comprise the steps of calculating a probability that a pixel of the selected image belongs to a foreground-relevant object according to a semantic segmentation algorithm, and assigning the pixel to the background pixel set if the probability that the pixel belongs to a foreground-relevant object does not exceed a first predetermined threshold, assigning the pixel to the foreground pixel set if the probability that the pixel belongs to a foreground-relevant object exceeds the first predetermined threshold and a difference between the probability that the pixel belongs to a foreground-relevant object and a baseline probability for the pixel equals or exceeds a second predetermined threshold.

Humans can easily delineate relevant moving objects with a high precision because they incorporate knowledge from the semantic level: they know what a car is, recognize shadows, distinguish between object motion and camera motion, etc. The purpose of semantic segmentation (also known as scene labeling or scene parsing) is to provide such information by labeling each pixel of an image with the class of its enclosing object or region. The task is difficult and requires the simultaneous detection, localization, and segmentation of semantic objects and regions. However, the advent of deep neural networks within the computer vision community and the access to large labeled training datasets have dramatically improved the performance of semantic segmentation algorithms, as described by J. Long, E. Shelhamer, and T. Darrell in "Fully convolutional networks for semantic segmentation", IEEE Int. Conf. Comput. Vision and Pattern Recogn. (CVPR), June 2015, pp. 3431-3440, by S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. Torr in "Conditional random fields as recurrent neural networks", IEEE Int. Conf. Comput. Vision (ICCV), December 2015, pp. 1529-1537, and by

H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," CoRR, vol. abs/1612.01105, December 2016. Semantic segmentation algorithms have thus begun to be used for specific computer vision tasks, such as optical flow estimation as described by L. Sevilla-Lara, D. Sun, V. Jampani, and M. J. Black in "Optical flow with semantic segmentation and localized layers", IEEE Int. Conf. Comput. Vision and Pattern Recogn. (CVPR), June 2016, pp. 3889-3898.

By leveraging information from a higher, semantic level, the method according to this first aspect of the disclosure can provide a more robust, reliable image segmentation into foreground and background than that provided by a background subtraction algorithm merely comparing low-level pixel values with a background model.

On one hand, the semantic level can thus be used to identify foreground-relevant objects, that is, objects belonging to semantic classes that can be expected to move, and thus belong to the foreground, and leverage this knowledge in the step of assigning the pixel to the background pixel set if the probability that the pixel belongs to a foreground-relevant object does not exceed a first predetermined threshold, so as to prevent false positives, that is, erroneously assigning pixels to the foreground pixel set due to, for example, dynamic backgrounds, ghosting, shadows and/or reflections, camera jitter, panning, tilting and/or zooming, bad weather, gradual or sudden lighting changes or background displacement, which usually affect the performances of conventional background subtraction algorithms.

On the other hand, the semantic level can also be used to identify whether the probability that a pixel belongs to such a foreground-relevant object is increased with respect to a baseline probability for that pixel, that may for instance correspond to a corresponding pixel in a semantic background model, in the step of assigning the pixel of the selected image to the foreground pixel set if the probability that the pixel belongs to a foreground-relevant object exceeds the first predetermined threshold and a difference between the probability that the pixel belongs to a foreground-relevant object and a baseline probability for the pixel equals or exceeds a second predetermined threshold, so as to prevent false negatives, that is, erroneously assigning pixels to the background, due to camouflage, i.e. when background and foreground share similar pixel values.

According to a second aspect of the present disclosure, the abovementioned method may further comprise a baseline updating step, wherein the baseline probability for the pixel is made equal to the probability that the pixel belongs to a foreground-relevant object calculated according to the semantic segmentation algorithm, if the pixel has been assigned to the background pixel set. Consequently, the baseline probability for the pixel can be updated for subsequent use with respect to corresponding pixels in other images using the information from the semantic level of this image. However, to avoid corrupting this baseline probability, for instance due to intermittent and slow-moving objects, a conservative updating strategy may be applied in which the baseline updating step is executed only randomly, according to a predetermined probability of execution, if the pixel has been assigned to the background pixel set.

According to a third aspect of the present disclosure, the method may further comprise a step of assigning the pixel to either the foreground pixel set or the background pixel set according to a background subtraction algorithm comparing a pixel value of the pixel with a pixel value of a corresponding pixel in a background model, and in particular a background subtraction algorithm based on at least one low-level

image feature, if the probability that the pixel belongs to a foreground-relevant object exceeds the first predetermined threshold and the difference between the probability that the pixel belongs to a foreground-relevant object and the baseline probability for the pixel is lower than the second predetermined threshold. Consequently, any pixel that the abovementioned steps fail to assign to either the foreground pixel set or the background pixel set on the basis of the semantic segmentation algorithm may be assigned using a known background subtraction algorithm comparing a pixel value of the pixel with a pixel value of a corresponding pixel in a background model. In particular, the pixel may belong to an image and the background model be based on at least another, related image, such as for instance a previous image in a chronological sequence of images including the image to which the pixel belongs.

Indeed, according to a fourth aspect of the present disclosure, the pixel may belong to an image of a chronological sequence of images, in which case the baseline probability for the pixel may have been initialized as equal to a probability that a corresponding pixel in an initial image of the plurality of related images belongs to a foreground-relevant object, calculated according to the semantic segmentation algorithm. Consequently, the semantic knowledge from this initial image can be leveraged in at least initially setting the baseline probabilities with which the probabilities of corresponding pixels in subsequent images belonging to foreground-relevant objects are compared when determining whether those pixels are to be assigned to the foreground.

According to a fifth aspect of the present invention, the first and second predetermined thresholds may have been predetermined so as to optimize an F score of the method on a test image series. The F score of a detection method may be defined as the harmonic mean between precision and recall, wherein the precision is a ratio of true positives to the sum of true positives and false positives and the recall is a ratio of true positives to the sum of true positives and false negatives. In the present context, pixels that are correctly assigned to the foreground can be considered as true positives, pixels that are incorrectly assigned to the foreground represent false positives, and pixels that are incorrectly assigned to the background represent false negatives. Consequently, predetermining the first and second predetermined thresholds so as to optimize the F score of the abovementioned method on a test image series can ensure a good compromise between precision p and recall r when the method is subsequently carried out on the selected image.

However, in an alternative sixth aspect of the present invention, the first and second predetermined thresholds may have been heuristically predetermined based on, respectively, a false foreground detection rate of the background subtraction algorithm and a true foreground detection rate of the background subtraction algorithm. It has indeed been found by the inventors that the first and second predetermined thresholds with which the F score of the method on test image series can be optimized are strongly correlated with, respectively, the false foreground detection rate and the true foreground detection rate of the background subtraction algorithm applied in this method. Consequently, if those rates are known from earlier tests of the background subtraction algorithm, it becomes possible to ensure a good compromise between precision and recall when the method is carried out on the selected image, even without carrying out a time- and resource-consuming optimization of the F score of the method applying both the background subtraction algorithm and the semantic segmentation algorithm.

According to a seventh aspect of the invention, a post-processing step has been added to the abovementioned method to address another challenge resulting from the combination of a semantic segmentation algorithm with a background subtraction algorithm. Indeed the method of assignment according to a semantic segmentation algorithm takes much more time than the method of assignment according to a background subtraction algorithm. The post-processing step is based on the fact that a true class (foreground or background) of a pixel is generally unchanged between two consecutive images of a series of images such as a chronological series of images in a video, as long as the object in that pixel remains static. One therefore assumes that if a correct assignment is enforced according to the first aspect of the invention, for a given pixel position, the same assignment should be enforced in that pixel position for the subsequent images if the semantic segmentation is too slow to calculate the probability for that pixel position of the subsequent images to belong to a foreground-relevant object.

The post-processing step, which may be part of a method called Asynchronous Semantic Post Processing (ASPP), will therefore memorize the information about the last enforced assignment for a given pixel position as well as the pixel value when the semantic segmentation algorithm has calculated the probability for a pixel to belong to a foreground-relevant object, and copy the last enforced assignment when the assignment rate of the semantic segmentation algorithm is too slow to calculate the probability for that pixel position of the selected image to belong to a foreground-relevant object, if the said assignment was enforced according to the first aspect of the invention and if the pixel value remains similar to a previously memorized pixel value for that pixel position. The information about the last enforced assignment for a given pixel position is stored in an assignment memory image, whereas the information about the pixel value is stored in a memory image.

Practically, it means that the method may further comprise a step of assigning a pixel of the selected image to the background pixel set or to the foreground pixel set even if the probability that the pixel belongs to a foreground-relevant object is not calculated with the semantic segmentation algorithm. This is performed by comparing the pixel value of that pixel of the selected image to the memorized pixel value for that position.

If the positive difference between the pixel value of the selected image and the memorized pixel value is lower than a third predetermined threshold and the pixel was assigned to the background pixel set according to the semantic probability calculated according to the first aspect of the present invention, then the pixel is assigned to the background pixel set.

Likewise, if the positive difference between the pixel value of the selected image and the memorized pixel value is lower than a fourth predetermined threshold and the pixel was assigned to the foreground pixel set according to the semantic probability calculated according to the first aspect of the present invention, then the pixel is assigned to the foreground pixel set. In other words, for both above-mentioned conditions, the last assignment enforced by the semantic segmentation is copied or repeated.

If none of both above-mentioned conditions are met, then it is preferable not to repeat a previous assignment and it is the background subtraction algorithm that determines the assignment of the pixel to the background pixel set or the foreground pixel set.

According to an eight aspect of the invention a semantic feedback step may further be added to the post processing step. Such semantic feedback step allows to upgrade the background subtraction algorithm itself.

As background subtraction algorithm, one may consider for example the well-known conservative background subtraction algorithm ViBe described by M. Van Droogenbroeck in "Source code for ViBe" at https://orbi.uliege.be/handle/2268/145853 or SuBSENSE described by P.-L. St-Charles at https://bitbucket.org/pierre-luc-st-charles/subsence.

As semantic feedback step, one may consider an assigned image as calculated according to the invention, to replace the assigned image provided by the background subtraction algorithm to update the background model of the background subtraction algorithm.

The present invention relates also to a data processing device programmed so as to carry out the image background recognition method of the invention; to a data storage medium comprising a set of instructions configured to be read by a data processing device to carry out an image background recognition method according to the invention; to a set of signals in magnetic, electromagnetic, electric and/or mechanical form, comprising a set of instructions for a data processing device to carry out an image background recognition method according to the invention; and/or to a process of transmitting, via magnetic, electromagnetic, electric and/or mechanical means, a set of instructions for a data processing device to carry out an image background recognition method according to the invention.

As "data storage medium" may be understood any physical medium capable of containing data readable by a reading device for at least a certain period of time. Examples of such data storage media are magnetic tapes and discs, optical discs (read-only as well as recordable or re-writable), logical circuit memories, such as read-only memory chips, random-access memory chips and flash memory chips, and even more exotic data storage media, such as chemical, biochemical or mechanical memories.

As "electromagnetic" any part of the electromagnetic spectrum is understood, from radio to UV and beyond, including microwave, infrared and visible light, in coherent (LASER, MASER) or incoherent form.

As "object" is understood any observable element of the real world, including animals and/or humans.

The above summary of some aspects of the invention is not intended to describe each disclosed embodiment or every implementation of the invention. In particular, selected features of any illustrative embodiment within this specification may be incorporated into an additional embodiment unless clearly stated to the contrary.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention may be more completely understood in consideration of the following detailed description of various embodiments in connection with the accompanying drawings, in which:

FIG. 1 is a flowchart illustrating a core routine of a method according to an aspect of the present disclosure;

FIG. 1A is a modified version of the flowchart of FIG. 1.

FIG. 2 is a functional scheme illustrating how the results of a semantic segmentation algorithm and a background subtraction algorithm are combined in the core routine of FIG. 1;

FIGS. 3A and 3B are graphs charting the positive correlations between the False Positive Rate FPR and True

Positive Rate TPR of the background subtraction algorithm and the optimum values for, respectively, a first predetermined threshold T_{BG} and a second predetermined threshold T_{FG} in the method of FIG. 2;

FIGS. 4 to 6 are graphs charting the improvement achieved by the method of FIG. 1 over a background subtraction algorithm used therein; and

FIG. 7 illustrates the outputs of a semantic segmentation algorithm, a background subtraction algorithm and a method combining both for various video sequences in difficult scenarios; and

FIGS. 8 and 9 illustrate potential embodiments of video systems applying the method of FIG. 1.

FIG. 10 is a functional scheme showing the ASPP method including the post-processing step according to the 7th aspect of the invention.

FIG. 11 is a possible timing diagram of ASPP when executed on a computer comprising a computer processing unit (CPU) and a graphical processing unit (GPU).

FIG. 12 is a graph charting the overall F score obtained with SBS and ASPP.

FIG. 13 are graphs charting the improvement achieved by the SBS and ASPP methods over a background subtraction algorithm used therein.

FIG. 14 is a graph charting the evolution of the optimum value of the third predetermined threshold τ_A and fourth predetermined threshold τ_B of ASPP with respect to the temporal sub-sampling factor.

FIG. 15 is a graph comparing the overall F score obtained by using the post-processed image showing the assignment to the foreground or background pixel sets, to update the background subtraction model.

While the invention is amenable to various modifications and alternative forms, specifics thereof have been shown by way of example in the drawings and will be described in detail. It should be understood, however, that the intention is not to limit aspects of the invention to the particular embodiments described. On the contrary, the intention is to cover all modifications, equivalents, and alternatives falling within the scope of the invention.

DETAILED DESCRIPTION

For the following defined terms, these definitions shall be applied, unless a different definition is given in the claims or elsewhere in this specification.

All numeric values are herein assumed to be preceded by the term "about", whether or not explicitly indicated. The term "about" generally refers to a range of numbers that one of skill in the art would consider equivalent to the recited value (i.e. having the same function or result). In many instances, the term "about" may be indicative as including numbers that are rounded to the nearest significant figure.

As used in this specification and the appended claims, the singular forms "a", "an", and "the" include plural referents unless the content clearly dictates otherwise. As used in this specification and the appended claims, the term "or" is generally employed in its sense including "and/or" unless the content clearly dictates otherwise.

The term "probability" is to be understood in a sense of probability estimate.

The term "assigned image" means an image wherein each pixel value is a class (background or foreground) produced by a background subtraction algorithm or assigned according to the first, or the third, or the seventh aspect of the invention.

The term "assignment memory image" means the image storing the assignments done according to the first or third or used in the seventh aspect of the invention.

The term "memory image" means an image wherein each pixel value has been copied or repeated from a previously selected image of the series of images according to the seventh aspect of the invention.

The term "SBS" means Semantic Background Subtraction. It is particularly used in the present invention in its first to fourth aspects.

The term "temporal sub-sampling" means skipping calculation of the semantic segmentation algorithm for some images of the series of images.

The term "spatial sub-sampling" means skipping calculation of the semantic segmentation algorithm for some regions of the selected image of the series of images.

The term "semantic image" means the image containing the probabilities according to the first aspect of the invention.

The term "conservative" for a background subtraction algorithm means that the background model is updated only if the pixel is assigned to the background pixel set.

The following detailed description should be read with reference to the drawings in which similar elements in different drawings are numbered the same. The detailed description and the drawings, which are not necessarily to scale, depict illustrative embodiments and are not intended to limit the scope of the invention. The illustrative embodiments depicted are intended only as exemplary. Selected features of any illustrative embodiment may be incorporated into an additional embodiment unless clearly stated to the contrary.

In a set of images that may have been, for example, successively captured by an imaging device at times t following a time series, each image may be formed by a plurality of pixels, each single pixel in an image having a dedicated pixel position x and a pixel value $\vec{T}(x)$. For ease of understanding, in the accompanying drawings, the pixel position x is shown as two-dimensional, but it could have any number of dimensions. For 3D images, for instance, the pixel position x may have three dimensions. The pixel value $\vec{T}(x)$ in the illustrated embodiment is a three-dimensional vector, in the form of RGB- or HSV-triplets for obtaining a polychromatic image. In alternative embodiments, it could however have any other number of dimensions.

A semantic segmentation algorithm may be applied to each image in order to calculate, for each pixel position x and time t, a real-valued vector $\vec{v}_t(x)=[v_t^1(x), v_t^2(x), \ldots, v_t^N(x)]$, where $v_t^i(x)$ denotes a score for each class c_i of a set $C=\{c_1, c_2, \ldots, c_N\}$ of N disjoint classes of objects. A set of probabilities $p_i(x \in c_i)$ that the pixel at pixel position x and time t belongs to each class c_i of the set C may be calculated by applying a softmax function to the vector $\vec{v}_t(x)$.

Among the N disjoint classes of objects of set C, a subset R may correspond to foreground-relevant objects, that is, objects relevant to motion detection. For instance, if the images relate to a street scene, these foreground-relevant objects may comprise potentially mobile objects like vehicles, people and animals, but not typically immobile objects like trees or buildings. Using the semantic segmentation algorithm it is thus possible to calculate an aggregated probability

$$p_{S,t}(x) = p_t(x \in R) = \sum_{c_i \in R} p_t(x \in c_i)$$

that the pixel at pixel position x and time t belongs to a foreground-relevant object. It may be possible to consider different subsets R, possibly with different numbers of disjoint classes of foreground-relevant objects, for different areas of an image. For instance, when the image shows both a road and a walking path, the subset R may include just people and animals as foreground-relevant object classes in the area of the walking path, but also vehicles in the area of the road.

This probability $p_{S,t}(x)$ that the pixel at pixel position x and time t belongs to a foreground-relevant object according to the semantic segmentation algorithm can be used in a method for assigning pixels to foreground and background pixel sets in each image of the set of images. FIG. 1 shows a flowchart illustrating a core routine of this method, with FIG. 1A showing a modified version thereof wherein the pixel at pixel position x and time t is assigned to either the foreground pixel set or the background pixel set. In a first step S100, S101 the probability $p_{S,t}(x)$ that the pixel at pixel position x and time t belongs to a foreground-relevant object is calculated using the semantic segmentation algorithm. In a second step S200, S201, it is determined whether this probability $p_{S,t}(x)$ is lower than or equal to a first predetermined threshold T_{BG}. If the result of this comparison is positive, and it is thus determined that the probability $p_{S,t}(x)$ that the pixel at pixel position x and time t belongs to a foreground-relevant object according to the semantic segmentation algorithm does not exceed the first predetermined threshold T_{BG}, it is considered unlikely that the pixel at pixel position x and time t belongs to a potentially mobile object, and the pixel at pixel position x and time t is thus assigned to the background in step S300. Using a binary variable $D \in [BG,FG]$ wherein the value BG indicates a background pixel and the value FG indicates a foreground pixel, this can be expressed as a first rule:

$$S_t^{BG}(x) \leq T_{BG} \rightarrow D_t(x) = BG$$

wherein $S_t^{BG}(x)$ denotes a signal that equals the probability $p_{S,t}(x)$, and $D_t^{(x)}$ denotes the value of the binary variable D for the pixel at pixel position x and time t. This first rule provides a simple way to address the challenges of illumination changes, dynamic backgrounds, ghosts and strong shadows, which severely affect the performances of conventional background subtraction algorithms by erroneously assigning background pixels to the foreground pixel set.

On the other hand, if in step S200, S201 it is determined that the probability $p_{S,t}(x)$ is not lower than or equal to a first predetermined threshold T_{BG}, in the next step S400, S401 it is determined whether a difference $S_t^{FG}(x)=p_{S,t}(x)-M_t(x)$ is at least equal to a second predetermined threshold T_{FG}, wherein $M_t(x)$ denotes a baseline probability for pixel position x and time t. This baseline probability $M_t(x)$ corresponds to a semantic model of the background for pixel position x and time t. It may have been initialized as equal to the probability $p_{S,0}(x)$ that a corresponding pixel at pixel position x and time, that is, in an initial image of the set of related images, belongs to a foreground-relevant object according to the semantic segmentation algorithm. It may then have been updated according to the following update strategy at each subsequent time step:

$$D_t(x)=FG \rightarrow M_{t+1}(x)=M_t(x)$$

$$D_t(x)=BG \rightarrow_\alpha M_{t+1}(x)=p_{S,t}(x) \rightarrow_{1-\alpha} M_{t+1}(x)=M_t(x)$$

wherein $\rightarrow_\alpha$ denotes a predetermined probability α of execution, which may be set, for example, to 0.00024. Therefore, the value of the baseline probability $M_{t+1}(x)$ for

pixel position x and the next time step t+1 is maintained equal to the baseline probability $M_t(x)$ for a corresponding pixel at time step t, and only updated randomly, according to the predetermined probability of execution P, with the value of the probability $p_{S,t}(x)$, if $D_t(x)$ BG, that is, if the pixel at pixel position x and time t has been assigned to the background pixel set.

The random determination, with predetermined probability α of execution, of whether the baseline probability $M_t(x)$ for pixel position x is to be updated, may be carried out using a random number generator. However, since such random numbers cannot be provided by a deterministic computer, a pseudorandom number generator may be used instead with properties similar to those of a true random number generator. Another alternative is the use of a large look-up list of previously generated random or pseudorandom numbers.

If the result of the comparison at step S400, S401 is positive and it is thus determined that the difference $S_t^{FG}(x)$ is indeed equal to or higher than this second predetermined threshold T_{FG}, it is considered that there has been a significant increase in the probability $p_{S,t}(x)$ for the pixel at pixel position x and time t with respect to that to be expected according to the semantic model, and in the next step S500 the pixel at pixel position x and time t is thus assigned to the foreground pixel set. This can be expressed as a second rule:

$$S_t^{FG}(x) \geq T_{FG} \rightarrow D_t(x) = FG$$

This second rule can prevent to a large extent that foreground pixels that are camouflaged, that is, that have similar pixel values to the background, be erroneously assigned to the background, which is also a challenge for conventional background subtraction algorithms.

However, semantic segmentation alone may not suffice to distinguish between foreground and background, for instance in the case in which a foreground-relevant object (e.g. a moving car) moves in front of a stationary object of the same, foreground-relevant object class (e.g. a parked car). Since both objects belong to the same foreground-relevant object class, the probability $p_{S,t}(x)$ will not significantly increase as the moving object moves in front of the stationary object at pixel position x and time t.

To address such a situation, if the result of the comparison at step S400, S401 is negative, that is, if the probability $S_t^{BG}(x)$ exceeds the first predetermined threshold T_{BG} and the difference $S_t^{FG}(x)$ is lower than the second predetermined threshold T_{FG} a third rule is applied in the next step S600, assigning the pixel at pixel position x and time t to either the foreground pixel set or the background pixel set according to a conventional background subtraction algorithm comparing a pixel value $\vec{T}(x)$ of the pixel at pixel position x and time t with a pixel value of a corresponding pixel in a background model based on at least another image of the plurality of related images. This can be expressed as a third rule:

$$D_t(x) B_t(x)$$

wherein $B_t(x) \in [BG, FG]$ denotes the result from the background subtraction algorithm.

FIG. 2 thus illustrates how the three signals $S_t^{BG}(x)$, $S_t^{FG}(x)$ and $B_t(x)$ can be obtained and applied in combination, using the abovementioned three rules, for foreground and background detection. How these signals are combined can also be summarized with the following table:

TABLE 1

Foreground and background detection according to the three rules of the method			
$B_t(x)$	$S_t^{BG}(x) \leq \tau_{BG}$	$S_t^{FG}(x) \geq \tau_{FG}$	$D_t(x)$
BG	false	false	BG
BG	false	true	FG
BG	true	false	BG
BG	true	true	X
FG	false	false	FG
FG	false	true	FG
FG	true	false	BG
FG	true	true	X

If the first predetermined threshold T_{BG} is set lower than the second predetermined threshold T_{FG}, the two situations denoted with "X" on Table 1 above cannot effectively be encountered.

Because the first rule only assigns pixels to the background pixel set, raising the first predetermined threshold T_{BG} so that the first rule is applied more frequently can only decrease the True Positive Rate TPR, that is the rate at which pixels are correctly assigned to the foreground, and the False Positive Rate FPR, that is the rate at which pixels are erroneously assigned to the foreground pixel set. On the other hand, because the second rule only assigns pixels to the foreground pixel set, decreasing the second predetermined threshold T_{FG} so that the second rule is applied more frequently can only increase the True Positive Rate TPR and the False Positive Rate FPR. Ideally, the first predetermined threshold T_{BG} and second predetermined threshold T_{FG} are thus to be set at the level that achieves the best compromise between the highest possible True Positive Rate TPR and the lowest possible False Positive Rate FPR.

One first alternative for setting the first predetermined threshold T_{BG} and second predetermined threshold T_{FG} is to perform tests on test image sequences using the abovementioned method with different values for the first predetermined threshold T_{BG} and second predetermined threshold T_{FG}, and select the duple of values for these thresholds resulting, for given background subtraction and semantic segmentation algorithms, in the best overall F score, that is, the highest harmonic mean between precision and recall, wherein the precision is a ratio of true positives (instances of pixels correctly assigned to the foreground pixel set) to the sum of true positives and false positives (instances of pixels erroneously assigned to the foreground pixel set) and the recall is a ratio of true positives to the sum of true positives and false negatives (instances of pixels erroneously assigned to the background pixel set). This can be performed as a grid search optimization.

The inventors have carried out such tests on 53 video sequences, organized in 11 categories, of the CDNet dataset presented by Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar in "CDnet 2014: An expanded change detection benchmark dataset", IEEE Int. Conf. Comput. Vision and Pattern Recogn. Workshops (CVPRW), pages 393-400, Columbus, Ohio, USA, June 2014, applying the presently disclosed method using 34 different background subtraction algorithms and, as semantic segmentation algorithm, the deep architecture PSPNet disclosed by H. Zhao, 3. Shi X. Qi, X. Wang and J. Jia in "Pyramid scene parsing network", CoRR, vol. abs/1612.01105, trained on the ADE20K dataset presented by B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso and A. Torralba in "Semantic understanding of scenes through the ADE20K dataset", CoRR, vol. abs/1608.05442, August 2016, to extract seman-

tics, using the PSPNet50 ADE20K model made publicly available by H. Zhao, running at approximately 7 fps at a 473×473 pixel image resolution on an NVIDIA® GeForce® GTX Titan X GPU. The last layer of this PSPNet architecture assigns to each pixel a score for each class c_i of a set $C=\{c_1, c_2, \ldots, c_N\}$ of N=150 disjoint object classes. In these tests, the selected subset of foreground-relevant object classes is R={person, car, cushion, box, book, boat, bus, truck, bottle, van, bag, bicycle}, corresponding to the semantics of CDNet foreground objects.

During these tests, it was found that there is a close correlation between the False Positive Rate FPR and True Positive Rate TPR of the background subtraction algorithm used in the third rule of the abovementioned method and, respectively, the first predetermined threshold T_{BG} and second predetermined threshold T_{FG} that achieve the best overall F score when applied in the first and second rules of the same method, as shown in FIGS. 3A and 3B.

Consequently, a second alternative approach is to heuristically set the first predetermined threshold T_{BG} and second predetermined threshold T_{FG} based on, respectively, the False Positive Rate FPR and True Positive Rate TPR of the background subtraction algorithm to be used in the third rule of the method. For instance, the first predetermined threshold T_{BG} may be set as half the False Positive Rate FPR of the background subtraction algorithm, and the second predetermined threshold T_{FG} as equal to the True Positive Rate TPR of the background subtraction algorithm. Since the background subtraction algorithm should, by definition, perform better than a random classifier, its False Positive Rate FPR should be lower than its True Positive Rate TPR, thus ensuring that the first predetermined threshold T_{BG} is also lower than the second predetermined threshold T_{FG}.

According to a third alternative approach, the first predetermined threshold T_{BG} and second predetermined threshold T_{FG} may be set to default values, corresponding for example to the arithmetic mean of the values for these thresholds resulting in the best overall F score for each of the best-performing five background subtraction algorithms in the 2014 CDNet ranking, with the same semantic segmentation algorithm.

Each one of these three alternative approaches has been tested and found to provide very significant improvements over the results of the underlying background subtraction algorithm on its own. FIG. 4 illustrates this improvement, defined as one minus the error rate of the method combining background subtraction with semantic segmentation divided by the error rate of the background subtraction algorithm on its own, for each one of these three approaches. More specifically, FIG. 4 illustrates the mean improvement, measured on the overall CDNet 2014 dataset, both for the entire set of 34 background subtraction algorithms, and for only the 5 best-performing background subtraction algorithms. As can be seen on this figure, the first approach offers a very significant improvement, even over the background subtraction algorithms that already performed best, and this improvement is hardly decreased with the second and third alternative approaches.

FIG. 5 illustrates the improvement, with respect to each background subtraction algorithm of the abovementioned set of 34 different background subtraction algorithms, in terms of change in the mean True Positive Rate TPR and False Positive Rate FPR. As can be seen there, the present method tends to reduce significantly the False Positive Rate FPR, while simultaneously increasing the True Positive Rate TPR.

FIG. 6 illustrates the mean improvement, both for all 34 different background subtraction algorithms and for the 5 best-performing, per category of video sequence, in the 11 categories of the CDNet 2014 dataset: "Baseline", "Dynamic background", "Camera jitter", "Intermittent object motion", "Shadow", "Thermal", "Bad weather", "Low framerate", "Night videos", "Pan-Tilt-Zoom Camera" and "Turbulence". Particularly good improvements can be observed for the "Baseline", "Dynamic background", "Shadow" and "Bad weather" categories. With respect to the "Thermal" and "Night videos" categories, it must be noted that the ADE20K dataset used to teach the semantic segmentation algorithm did not include images of these types, which may explain the less good results for those categories.

FIG. 7 illustrates the benefits of the method according to the present disclosure in four different scenarios of real-world surveillance tasks. From left to right, the four columns correspond, respectively, to scenarios with dynamic background, risk of ghosts, strong shadows, and camouflage effects. From the top down, the five rows illustrate a frame of the corresponding video sequence, the probability $p_{S,t}(x)$ for each pixel, the output of the IUTIS-5 background subtraction algorithm described by S. Bianco, G. Ciocca and R. Schettini in "How far can you get by combining change detection algorithms?", CoRR, vol. abs/1505.02921, 2015, the output of the presently-disclosed method, applying the IUTIS-5 background subtraction algorithm in its third rule, and the ground truth. As can be seen, with respect to the IUTIS-5 background subtraction algorithm on its own, the presently-disclosed method greatly reduces the number of false positive foreground pixel detections caused by dynamic backgrounds, ghosts and strong shadows, while at the same time mitigating camouflage effects.

FIG. 10 illustrates the post-processing step included in the Asynchronous Semantic Post-Processing (ASPP) that is described in details hereafter.

Let us first focus on a substitute for the first rule according to the first aspect of the invention, denoted rule A hereafter, that replaces the first rule in the absence of semantics. If the first rule was previously activated in pixel location x while the current pixel color has remained similar, then $D_t(x)$ should be assigned to the background pixel set. To enable this mechanism, we have to store, in an assignment memory image denoted by R, if the first rule of semantic background subtraction (SBS) is activated in position x; this is indicated by $R(x) \leftarrow 1$. Simultaneously, we memorize the value of that pixel in a memory image, whose value at pixel position x is denoted by $\vec{C}(x)$. With these components, rule A becomes:

if $R(x)=1$ and dist$\{\vec{C}(x), \vec{I}(x)\} \leq \tau_A$ then $D_t(x)$ BG where τ_A is a third predetermined threshold applied on the Manhattan (or Euclidean) distance between the pixel value stored in the memory image $\vec{C}(x)$ and the pixel value $\vec{I}(x)$.

Likewise, we can replace the second rule according to the first aspect by rule B in the absence of semantics. When the second rule is activated in position x, this assignment is stored in the assignment memory image (this is indicated by $R(x) \leftarrow 2$, and the value of the pixel is stored in the memory image at that position. Rule B thus becomes:

if $R(x)=2$ and dist$\{\vec{C}(x), \vec{I}(x)\} \leq \tau_B$ then $D_t(x)$ FG where τ_B is a fourth predetermined threshold. Again, when neither rule A nor rule B are activated, the background subtraction algorithm is used to assign the pixel to the foreground pixel set or the background pixel set.

The updates of the assignment memory image and memory image are detailed in Algorithm 1.

Algorithm 1 Pseudo-Code for Pixels with Semantics. The Assignment Memory Image and Memory Image are Updated During the Application of SBS.

Require: $T(x)$ is the selected image of the images series (at time t)
Require: R is initialized with zero values
1:      for all pixel positions x with a calculated aggregated semantic probability do
2:          $D_t(x) \leftarrow$ apply SBS in x
3:          if the first rule is activated then
4:              $R(x) \leftarrow 1$
5:              $\bar{C}(x) \leftarrow T(x)$
6:          else if the second rule is activated then
7:              $R(x) \leftarrow 2$
8:              $\bar{C}(x) \leftarrow T(x)$
9:          else
10:             $R(x) \leftarrow 0$
11:         end if
12:     end for

Algorithm 1 is an add-on for SBS that memorizes assignments and values based on semantics. The second component of ASPP, described in Algorithm 2, is the application of rule A, rule B, or a fallback; when no semantics is computed.

Algorithm 2 Pseudo-Code of ASPP for Pixels without Semantics. Rule A and Rule B are Applied.

Require: $T(x)$ is the selected image of the images series (at time t))
1:      for all pixel positions x that do not have a calculated aggregated semantic probability do
2:          if $R(x)= 1$ then
3:              if $dist(\bar{C}(x),T(x)) \le \tau_A$ then
4:                  $D_t(x) \leftarrow BG$
5:              end if
6:          else if $R(x)= 2$ then
7:              if $dist(\bar{C}(x),T(x)) \le \tau_B$ then
8:                  $D_t(x) \leftarrow FG$
9:              end if
10:         else
11:             $D_t(x) \leftarrow B_t(x)$
12:         end if
13:     end for

Note that the two pseudo-codes, which are defined for each pixel, could be applied within the same image of the series of images such as in a video if the semantics was only calculated for a specific part of the selected image. It is therefore straightforward to adapt the method from a temporal sub-sampling to a spatial sub-sampling or a combination of both. However, a typical setup is that semantics is calculated for the whole image and is skipped for a few next images in a series of images at a regular basis. It is convenient to evaluate ASPP for this temporal sub-sampling since it has a unique implementation, while spatial sub-sampling can involve complex strategies for choosing parts where to calculate the semantics and is in any event application dependent. The ASPP method is illustrated in FIG. 10 for a temporal sub-sampling.

FIG. 10 is a synthetic flowchart of the ASPP method, extending SBS, capable to post-process the assigned image provided by a background subtraction algorithm even when semantics is not computed for all images of a series of images.

FIG. 11 illustrates a possible timing diagram of the ASPP method.

The ASPP method introduces a small computational overhead (a difference is calculated in terms of a distance for some pixels) and memory increase (an assignment memory image and a memory image are memorized) compared to SBS only. The practical benefits of ASPP can be visualized

on a detailed timing diagram of its components (see FIG. 11). For a formal discussion, we use the following notations:

I_t, S_t, B_t, D_t, respectively denote the selected image, semantic image, assigned image produced by a background subtraction algorithm, or assigned images produced by SBS or ASPP, indexed by a time index t.

δ_I represent the time between two consecutive images of the image series.

Δ_S, Δ_B, Δ_D are the times needed to calculate the semantic images, the assigned image produced by a background subtraction algorithm, and to apply the post-processing (SBS or ASPP), respectively. These times are supposed constant for the following.

Semantics is calculated on a GPU, whereas the background subtraction algorithm and post-processing algorithm are executed on a single threaded CPU hardware. Also, the assignment rate of the semantic segmentation algorithm is supposed to be smaller than that of the background subtraction algorithm; that is $\Delta s > \Delta B$.

We now examine two different scenarios. The first scenario is that of a real-time background subtraction algorithm (which implies that $\Delta_B \le \delta_I$) satisfying the condition $\Delta_B + \Delta_D \le \delta_I$. This scenario, illustrated in FIG. 11, can be obtained with the ViBe background subtraction algorithm for example. On the timing diagram, it can be seen that the assignment rate of ASPP is then equal to the assignment rate of the background subtraction algorithm with a time delay corresponding approximately to Δ_S.

In a second scenario, the assignment rate of the background subtraction algorithm is too slow to accommodate to real time with ASPP. It means that $\Delta_B + \Delta_D > \delta_I$. In this case, the assignment rate of ASPP is mainly dictated by Δ_B, since $\Delta_B \square \Delta_D$ ($\Delta_B >> \Delta_D$), just like when we do not have any post-processing mechanism. It is a scenario that follows the current trend to produce better background subtraction algorithms at the price of more complexity and slower assignment rates. Indeed, according to experiments led by the inventors, the top unsupervised background subtraction algorithms ranked on the CDNet web site (see http://change-detection.net) are not real time.

Finally, note that both SBS and ASPP do not impact the behavior of the background subtraction algorithm, which means that the internal background model of the background subtraction algorithm is not improved by semantics. This motivates the introduction of an additional feedback mechanism described hereinafter.

The post-processing SBS and ASP are designed to improve the quality of the final assigned image but they do not affect the assignments done by the background subtraction algorithm itself.

The backbone of many background subtraction algorithms is composed of three main parts. First, a background model of a scene is kept in a memory, for instance in the form of pixel values or other types of features. Second, the pixels of a selected image of the series of images are compared to this model, and consequently the pixels are assigned to the background pixel set or to the foreground pixel set. Third, the background model is updated to account for changes in the background of the scene over time.

According to the eighth aspect of this invention, we have added to ASPP a feedback mechanism for algorithms whose update policy is conservative. For those algorithms, the background model in pixel position x may be updated if $B_t(x) = BG$, but it is always left unchanged if $B_t(x) = FG$ which prevents the background model from being corrupted with foreground features. Therefore, it can be said that the pixel values of B_t determine if an update takes place. In other

words, the pixels of the B_t image assigned to the background pixel set determine the position of pixels whose background models are updated. As the assigned image D_t, which is post-processed by ASPP, is often an improved version of B_t, we can advantageously replace B_t by D_t to determine the positions of the pixels whose background model is updated. This introduces a semantic feedback from the assigned image and improves the background model of the background subtraction algorithm and consequently the next assigned image, whether or not semantics is computed.

Another possible refinement consists in adapting the update rate of the background model according to the assigned image produced by ASPP. More specifically, if $B_t(x)=FG$ and $D_t(x)=BG$, one could assume that the background model in pixel x is inadequate and, consequently, increase the update rate in that pixel. Tests performed on the well-known background subtraction algorithm ViBe showed that the performances are improved with this refinement.

Compared to the timing diagram of ASPP presented in FIG. 11, one has to include times related to a feedback using the assigned image D_t. This has two effects. First, the time to replace B_t by D_t for background model update, denoted Δ_F, is added to the time needed for calculations of the background subtraction algorithm Δ_B. This time is negligible, since it can be implemented as a simple pointer replacement. Second, if the post-processing is computed on a different CPU than the one used for the background subtraction algorithm, then one has to wait for the post-processing to finish before beginning the background subtraction of the next image of the series of images.

The inventors have evaluated the performance of the ASPP post-processing and have compared its performance to those of the original background subtraction algorithm and those of the SBS method applied on images with semantics (images with no semantic segmentation are not post-processed by SBS). These performance evaluations have been performed on the CDNet 2014 dataset, composed of 53 video sequences taken in various environmental conditions such as bad weather, dynamic backgrounds and night conditions, as well as different video acquisition conditions, such as PTZ and low frame rate cameras. This challenging dataset is largely employed within the background subtraction community and currently serves as the reference dataset to compare state-of-the-art background subtraction techniques.

The inventors have compared the performance on this dataset according to the overall F score (calculated according to the rules of CDNet 2014), which is one of the most widely used performance indicators for this dataset. For each video, F is computed as explained abovementioned.

For dealing with missing semantics, since the possibilities to combine spatial and temporal sampling schemes are endless, the inventors have restricted the performance evaluation to the case of a temporal sub-sampling of one semantic image per X original images of images series; this sub-sampling factor is referred to as X:1 hereafter. In other scenarios, semantics could be obtained at a variable rate or for some variable regions of interest, or even a mix of these sub-sampling schemes.

The four predetermined thresholds are chosen as follows. For each background subtraction algorithm, one first optimizes the first and second predetermined thresholds, respectively τ^{BG} and τ_{FG}, with a grid search to maximize the overall F score. Then, the optimal thresholds, denoted by τ_{BG}^* and τ_{FG}^*, are frozen and one optimizes the third and fourth predetermined thresholds, τ_A and τ_B of ASPP by a second grid search for each pair (background subtraction

algorithm, X:1), to maximize the overall F score once again. Such methodology allows a fair comparison between SBS and ASPP as the two techniques use the same common parameters τ_{BG}^* and τ_{FG}^* and the ASPP method is judged on its ability to improve SBS when SBS behaves optimally.

For comparison purposes, the assigned images of the background subtraction algorithms are taken directly from the CDNet 2014 website.

A comparison of the performances obtained with SBS and ASPP for four state-of-the-art background subtraction algorithms and for different sub-sampling factors is provided in FIG. 12. For the evaluation, the inventors considered two possibilities for comparing the application of SBS for images with no calculated semantics. For images missing semantics, it is not possible to apply the SBS directly. Therefore, a first possibility consists in copying B_t in D_t for images with no semantics. A second possibility consists to copy the semantic probabilities of a previously calculated semantic segmentation and to apply the assignment rules according to the first aspect of this invention, for images with no semantics.

The performances of ASPP decrease much more slowly than those of the two variants of SBS capable to handle images with no semantics, with respect to the decrease of the semantic image rate and, therefore, are much closer to those of the ideal case (SBS with all semantic images computed, that is SBS 1:1), meaning that ASPP provides better assigned images for images without semantics. It can be seen that, on average, ASPP with 1 image of semantics out of 25 images (ASPP 25:1) performs as well as SBS with 1 image of semantics out of 2 images (SBS 2:1).

FIG. 13 also compares the effects of SBS and ASPP on different background subtraction algorithms by looking at the performances in the mean ROC space of CDNet 2014 (ROC space where the false and true foreground rates are computed according to the rules of CDNet). Each point represents the performance of a background subtraction algorithm, whose assigned images are downloaded from the dataset website, and the end of the associated arrow indicates the performance after application of the post-processing method for a temporal sub-sampling factor of 5:1. This choice of sub-sampling rate is motivated by the fact that it is the image rate at which PSPNet can produce the semantic images on a NVIDIA® GeForce® GTX Titan X GPU. The inventors observed that, for images without semantics at a sub-sampling factor of 5:1, SBS improves the performances, but only marginally, whereas ASPP moves the performances much closer to the oracle (upper left corner).

FIG. 14 shows the evolution of the optimal parameters value of the third predetermined threshold τ_A and fourth predetermined threshold τ_B with respect to the temporal sub-sampling factor (in the case of PAWCS). The optimal value decreases with the sub-sampling factor, implying that rule A and rule B should be activated less frequently for lower semantic image rates, as a consequence of the presence of more outdated values in the memory image.

The inventors have also evaluated the performance of the semantic feedback step according to the eighth aspect of the invention. In this scenario, the assigned image D_t obtained from ASPP can be used in a feedback loop in order to improve the background model. It is valuable to improve the assignments of the background subtraction algorithm when no semantic frames are computed. The inventors have analyzed the results that were obtained for D_t in ASPP incorporating a semantic feedback loop.

The inventors performed tests for two well-known conservative background subtraction algorithms ViBe and SuB-

SENSE, using the code made available by the inventors of these algorithms, at the following address https://orbi.u-liege.be/handle/2268/145853 for ViBe and at the following address https://bitbucket.org/pierre-luc-st-charles/subsence for SuBSENSE. Let us note that the performances for SuBSENSE are lower than the ones reported in FIG. 12 as default parameters were used on each video.

FIG. 15 (left column) charts the overall F score on the CDNet 2014 dataset for methods according to the first, seventh and eighth (feedback) aspects of the invention for the ViBe and SuBSENSE background subtraction algorithms. Two main observations can be made. First, the SBS and ASPP both improve the performances even when the semantic frame rate is low. Also, ASPP always performs better. Second, including the feedback always improves the performances for both SBS and ASPP, and for both algorithms. In the case of ViBe, the performance is much better when the feedback is included. For SuBSENSE, the performance is also improved, but only marginally. This might be due to the fact that ViBe has a very straightforward way of computing the update of the background model while SuBSENSE uses varying internal parameters and heuristics, calculated adaptively. It is thus more difficult to interpret the impact of the semantic feedback on SuBSENSE than it is on ViBe.

The second investigated effect of the feedback is to what extent the feedback provides better assigned images to update the model of the background subtraction algorithm. For conservative algorithms such as the background subtraction algorithms ViBe and SuBSENSE, this means that, internally, the background model is built with better features and that can be evaluated using the assigned image provided by the background subtraction algorithm, that is B_t.

For that purpose, the inventors have compared the assigned image provided by the background subtraction algorithm B_t, for the semantic feedback when the background subtraction model is updated using D_t obtained by ASPP, instead of B_t. As shown in FIG. 15 (right column), using the semantic feedback always positively impacts the performance of the background subtraction algorithm. This means that the background model of the background subtraction algorithm is always improved compared to the original method which helps the algorithm to perform better assignments.

Finally, the inventors noted that the background subtraction algorithm ViBe, which is a real-time algorithm, combined with semantics provided at a real-time rate (about 1 out of 5 images of series of images) and with the feedback from ASPP has a mean performance of 0.744 which is about the same performance as the original SuBSENSE algorithm (0.746) that is not real-time. ASPP and the feedback are thus strategies that can help real-time algorithms to reach performances of background subtraction algorithms, such as SuBSENSE, while keeping the real-time constraint, which is a huge advantage in practice.

The presently disclosed method may be carried out with assistance of a data processing device, such as, for example, a programmable computer like the abovementioned NVIDIA® GeForce® GTX Titan X GPU, connected to an imaging device providing a video sequence of successive images. In such a case, the data processing device may receive instructions for carrying out this method using a data storage medium, or as signals in magnetic, electromagnetic, electric and/or mechanical form.

The presently disclosed method may, for example, be applied to video-surveillance, professional and/or consumer digital still and/or video cameras, computer and videogame

devices using image capture interfaces, satellite imaging and Earth observation, automatic image analysis and/or medical imaging systems or may be included in a smartphone.

FIG. 8 illustrates a possible application of the invention with an imaging device 1 in the particular form of a digital camera with an embedded data processing device 2 programmed to carry out the method of the invention. FIG. 9 illustrates another possible application of the invention with an imaging device 1 connected to a data processing device 2 programmed to carry out the method of the invention.

Those skilled in the art will recognize that the present invention may be manifested in a variety of forms other than the specific embodiments described and contemplated herein. Accordingly, departure in form and detail may be made without departing from the scope of the present invention as described in the appended claims.

The invention claimed is:

1. A method for assigning a pixel to one of a foreground pixel set and a background pixel set using asynchronous semantic post processing to improve motion detection by an imaging device, the pixel belonging to an image of a chronological sequence of images taken by the imaging device that includes background and foreground objects, the method comprising the steps of:

initializing a baseline probability for the pixel as equal to a probability that a corresponding pixel in an initial image of the chronological sequence of images belongs to a foreground relevant object, said probability being calculated according to a semantic segmentation algorithm;

calculating a probability that the pixel belongs to a foreground-relevant object according to a semantic segmentation algorithm; and

assigning the pixel to the background pixel set if the probability that the pixel belongs to a foreground-relevant object does not exceed a first predetermined threshold;

assigning the pixel to the foreground pixel set if the probability that the pixel belongs to a foreground-relevant object exceeds the first predetermined threshold and a difference between the probability that the pixel belongs to a foreground-relevant object and the baseline probability for the pixel equals or exceeds a second predetermined threshold;

using a background subtraction algorithm to assign the pixel to the foreground pixel set or to the background pixel set if the probability that the pixel belongs to the foreground-relevant object exceeds the first predetermined threshold and a difference between the probability that the pixel belongs to a foreground-relevant object and the baseline probability for the pixel is lower than a second predetermined threshold; and

said asynchronous semantic post processing assigning the pixel to the background pixel set or the foreground pixel set according to the last enforced assignment provided by the semantic segmentation algorithm for the given pixel position when an assignment rate of the semantic segmentation algorithm is too slow to calculate the probability that said pixel position belongs to a foreground-relevant object, said asynchronous semantic post processing reducing the rate of false positives, wherein false positives represent an erroneous assignment of pixels to the foreground pixel set due to one or more of dynamic backgrounds, ghosting, shadows, reflections, camera jitter, panning, tilting, zooming, inclement weather, gradual or sudden lighting changes

and background displacement in the sequence of images being taken by the imaging device, said asynchronous semantic post processing reducing a rate of false negatives, wherein false negatives represent an erroneous assignment of pixels to the background pixel set as can occur when at least some of the background and foreground objects in the image share similar pixel values.

2. The method as set forth in claim 1, wherein said step of assigning by said asynchronous semantic post processing includes the steps of:

comparing the pixel value to a memorized pixel value for said pixel position;

assigning the pixel to the background pixel set when the last enforced assignment provided by the semantic segmentation algorithm for the given pixel position has assigned the pixel to the background pixel set based upon having determined that the probability that the pixel belongs to a foreground-relevant object does not exceed the first predetermined threshold or that said probability has not been calculated, and upon determining that a positive difference between the pixel value and the memorized pixel value is lower than a third predetermined threshold.

3. The method as set forth in claim 1, wherein said step of assigning by said asynchronous semantic post processing includes the steps of:

comparing the pixel value to a memorized pixel value for said pixel position;

assigning the pixel to the foreground pixel set when the last enforced assignment provided by the semantic segmentation algorithm for the given pixel position has assigned the pixel to the foreground pixel set based upon having determined that the probability that the pixel belongs to a foreground-relevant object exceeds the first predetermined threshold, and that a difference between the probability that the pixel belongs to a foreground-relevant object and the baseline probability for the pixel equals or exceeds a second predetermined threshold; or that said probability has not been calculated and upon determining that a positive difference between the pixel value and the memorized pixel value is lower than a fourth predetermined threshold.

4. The method as set forth in claim 1, wherein said step of assigning by said asynchronous semantic post processing includes the steps of:

comparing the pixel value to a memorized pixel value for said pixel position and,

a) assigning the pixel to the background pixel set when the last enforced assignment provided by the semantic segmentation algorithm for the given pixel position has assigned the pixel to the background pixel set based upon having determined that the probability that the pixel belongs to a foreground-relevant object does not exceed the first predetermined threshold or that said probability has not been calculated, and upon determining that a positive difference between the pixel value and the memorized pixel value is lower than a third predetermined threshold;

b) assigning the pixel to the foreground pixel set when the last enforced assignment provided by the semantic segmentation algorithm for the given pixel position has assigned the pixel to the foreground pixel set based upon having determined that the probability that the pixel belongs to a foreground-relevant object exceeds the first predetermined threshold, and that a difference between the probability that the pixel belongs to foreground-relevant object and the baseline probability for

the pixel is equal or exceeds a second predetermined threshold or that said probability has not been calculated, and upon determining that a positive difference between the pixel value and the memorized pixel value is lower than a fourth predetermined threshold;

using a background subtraction algorithm to determine the pixel assignment if neither the conditions set forth in a) or b) are met.

5. The method as set forth in claim 2, further comprising a semantic feedback step that includes updating a background model of the background subtraction algorithm with the assignment calculated using said asynchronous semantic post processing.

6. The method as set forth in claim 1, further comprising the step of assigning the pixel to the foreground pixel set if the probability that the pixel belongs to a foreground-relevant object exceeds the first predetermined threshold and a difference between the probability that the pixel belongs to a foreground-relevant object and the baseline probability for the pixel equals or exceeds the second predetermined threshold so as to reduce a rate of false negatives, wherein false negatives represent an erroneous assignment of pixels to the background pixel set as can occur when at least some of the background and foreground objects in the image share similar pixel values.

7. An imaging device for taking a chronological sequence of images comprising a data processing device configured to assign a pixel from an image of said chronological sequence of images to one of a foreground pixel set and a background pixel set using asynchronous semantic post processing to improve motion detection by the imaging device by reducing the rate of false positives, in which false positives represent an erroneous assignment of pixels to the foreground pixel set due to one or more of dynamic backgrounds, ghosting, shadows, reflections, camera jitter, panning, tilting, zooming, inclement weather, gradual or sudden lighting changes and background displacement in the sequence of images being taken by the imaging device, said data processing device being programmed to:

initialize a baseline probability for the pixel as equal to a probability that a corresponding pixel in an initial image of the chronological sequence of images belongs to a foreground relevant object, said probability being calculated according to a semantic segmentation algorithm;

assign the pixel to the background pixel set if the probability that the pixel belongs to a foreground-relevant object does not exceed a first predetermined threshold;

assign the pixel to the foreground pixel set if the probability that the pixel belongs to a foreground-relevant object exceeds the first predetermined threshold and a difference between the probability that the pixel belongs to a foreground-relevant object and the baseline probability for the pixel equals or exceeds a second predetermined threshold;

use a background subtraction algorithm to assign the pixel to the foreground pixel set or to the background pixel set if the probability that the pixel belongs to the foreground-relevant object exceeds the first predetermined threshold and a difference between the probability that the pixel belongs to a foreground-relevant object and the baseline probability for the pixel is lower than a second predetermined threshold; and

use said asynchronous semantic post processing to assign the pixel to the foreground pixel set or the background pixel set according to the last enforced assignment provided by semantic segmentation algorithm for the

given pixel position when an assignment rate of the semantic segmentation algorithm is too slow to calculate the probability that said pixel position belongs to a foreground-relevant object.

8. The imaging device as set forth in claim 7, wherein said imaging device is a digital camera.

9. The imaging device as set forth in claim 7, wherein said data processing device is further programmed to use said asynchronous semantic post processing to compare the pixel value to a memorized pixel value for said pixel position, and assign the pixel to the background pixel set when the last enforced assignment provided by the semantic segmentation algorithm for the given pixel position has assigned the pixel to the background pixel set based upon having determined that the probability that the pixel belongs to a foreground-relevant object does not exceed the first predetermined threshold, or that said probability has not been calculated, and upon determining that a positive difference between the pixel value and the memorized pixel value is lower than a third predetermined threshold.

10. The imaging device as set forth in claim 9, wherein said data processing device is further programmed to use said asynchronous semantic post processing to compare the pixel value to a memorized pixel value for said pixel position, and assign the pixel to the foreground pixel set when the last enforced assignment provided by the semantic segmentation algorithm for the given pixel position has assigned the pixel to the foreground pixel set based upon having determined that the probability that the pixel belongs to a foreground-relevant object exceeds the first predetermined threshold, and that a difference between the probability that the pixel belongs to foreground-relevant object and the baseline probability for the pixel equals or exceeds a second predetermined threshold or that said probability has not been calculated and upon determining that a positive difference between the pixel value and the memorized pixel value is lower than a fourth predetermined threshold.

11. The imaging device as set forth in claim 10, wherein said imaging device is a digital camera.

12. The imaging device as set forth in claim 11, wherein said data processing device is further programmed to perform a semantic feedback step that includes updating a background model of the background subtraction algorithm with the assignment calculated using said asynchronous semantic post processing.

13. The imaging device as set forth in claim 11, wherein said data processing device is further programmed to use the background subtraction algorithm to determine the pixel assignment to the background pixel set or to the foreground pixel set when said asynchronous semantic post processing has determined that a previous assignment should not be repeated.

14. A non-transitory data storage medium comprising a set of instructions configured to be read by a data processing device, the set of instructions including a plurality of steps to be undertaken to assign a pixel, from an image of a chronological sequence of images taken by an imaging device, to one of a foreground pixel set and a background pixel set using asynchronous semantic post processing to improve motion detection by the imaging device, said plurality of steps in the set of instructions including initializing a baseline probability for the pixel as equal to a probability that a corresponding pixel in an initial image of the chronological sequence of images belongs to a foreground relevant object, said probability being calculated according to at a semantic segmentation algorithm; assigning the pixel to the background pixel set if the probability that the pixel belongs

to a foreground-relevant object does not exceed a first predetermined threshold; assigning the pixel to the foreground pixel set if the probability that the pixel belongs to a foreground-relevant object exceeds the first predetermined threshold and a difference between the probability that the pixel belongs to a foreground-relevant object and the baseline probability for the pixel equals or exceeds a second predetermined threshold;

using a background subtraction algorithm to assign the pixel to the foreground pixel set or to the background pixel set if the probability that the pixel belongs to the foreground-relevant object exceeds the first predetermined threshold and a difference between the probability that the pixel belongs to a foreground-relevant object and the baseline probability for the pixel is lower than a second predetermined threshold; using said asynchronous semantic post processing to assign the pixel to the foreground pixel set or the background pixel set according to a last enforced assignment provided by the semantic segmentation algorithm for a given pixel position when an assignment rate of the semantic segmentation algorithm is too slow to calculate the probability that said pixel position belongs to a foreground-relevant object;

wherein the steps performed by the data processing device according to the set of instructions reduce a rate of false positives which represent an erroneous assignment of pixels to the foreground pixel set due to one or more of dynamic backgrounds, ghosting, shadows, reflections, camera jitter, panning, tilting, zooming, inclement weather, gradual or sudden lighting changes and background displacement in the sequence of images taken by the imaging device.

15. The non-transitory data storage medium as set forth in claim 14, wherein said set of instructions further directs the data processing device to use said asynchronous semantic post processing to compare the pixel value to a memorized pixel value for said pixel position, and assign the pixel to the background pixel set when the last enforced assignment provided by the semantic segmentation algorithm for the given pixel position has assigned the pixel to the background pixel set based upon having determined that the probability that the pixel belongs to a foreground-relevant object does not exceed the first predetermined threshold or that said probability has not been calculated, and upon determining that a positive difference between the pixel value and the memorized pixel value is lower than a third predetermined threshold.

16. The non-transitory data storage medium as set forth in claim 14, wherein said set of instructions further directs the data processing device to use said asynchronous semantic post processing to compare the pixel value to a memorized pixel value for said pixel position, and assign the pixel to the foreground pixel set when the last enforced assignment provided by the semantic segmentation algorithm for the given pixel position has assigned the pixel to the foreground pixel set based upon having determined that the probability that the pixel belongs to a foreground-relevant object exceeds the first predetermined threshold, and that a difference between the probability that the pixel belongs to foreground-relevant object and the baseline probability for the pixel equals or exceeds a second predetermined threshold, or that said probability has not been calculated and upon determining that a positive difference between the pixel value and the memorized pixel value is lower than a fourth predetermined threshold.

* * * * *