# Neural Background Subtraction for Pan-Tilt-Zoom Cameras

Alessio Ferone, *Member, IEEE* and Lucia Maddalena, *Member, IEEE*

***Abstract*—We propose an extension of a neural-based background subtraction approach to moving object detection to the case of image sequences taken from Pan-Tilt-Zoom (PTZ) cameras. The background model automatically adapts in a self-organizing way to changes in the scene background. Background variations arising in a usual stationary camera setting, such as those due to gradual illumination changes, to waving trees, or to shadows cast by moving objects, are accurately handled by the neural self-organizing background model originally proposed for this type of setting. Handling of variations due to the PTZ camera movement is ensured by a novel registration mechanism that allows the neural background model to automatically compensate the eventual ego-motion, estimated at each time instant. Experimental results on several real image sequences and comparisons with seven state-of-the-art methods demonstrate the accuracy of the proposed approach.**

***Index Terms*—artificial neural network, background subtraction, motion detection, PTZ camera, self organization, video surveillance.**

## I. Introduction

AUTOMATED video surveillance using video analysis and understanding technology has become an important research topic in the area of computer vision. Within video understanding technology for surveillance use, moving object detection is known to be a significant and difficult research problem. Indeed, aside from the intrinsic usefulness of being able to segment video streams into moving and background components, moving object detection provides a focus of attention for recognition, classification, and activity analysis, making these later steps more efficient, since only moving pixels need to be considered [1].

Most cameras used in surveillance are fixed, allowing one to only look at one specific view of the surveilled area. For scenes taken from this type of cameras the most common and efficient approach to moving object detection is background subtraction, that consists in maintaining an up-to-date model of the fixed background and detecting moving objects as those that deviate from such model. Due to its pervasiveness in various contexts, background subtraction has been afforded by several researchers, and plenty of literature has been published (see surveys in [2]–[6]). Compared to other approaches, such as optical flow, this approach is computationally affordable for real-time applications, is independent on moving object velocity, and is not subject to the foreground aperture problem. Nevertheless, the background subtraction approach is highly sensitive to dynamic scene changes due to lighting and extraneous events, with the consequent need for a suitable adaptation of the background model [7], [8].

The growing proliferation of moving camera platforms, such as smart phones and robots, is highlighting the severe limits imposed by the assumption of camera stationarity [9], [10]. Many attempts have been done in order to solve the problem in the case of freely moving cameras [9], [11]–[14]. Recently, the progress in sensor technologies has lead to a growing dissemination of a specific type of moving cameras, the so-called Pan-Tilt-Zoom (PTZ) cameras, that can dynamically modify their field of view through the use of panning, tilting, and zooming (i.e., moving left and right, up and down, closer and farther away). Their functionality introduces new surveillance capabilities, such as increasing the resolution of moving targets and adapting the sensor coverage, thus enabling to focus the attention on automatically selected areas of interest. On the other hand, the PTZ camera movement has introduced new challenges. A first problem is represented by the fact that, due to the camera movement, even pixels belonging to static objects appear to move in the camera frame. Such an effect is called *ego-motion* and its estimation and compensation represents one of the main objectives of the active vision research area [15]. Extensive research carried out regarding moving object detection for PTZ cameras can be categorized into three approaches [16]:

- The *optical flow clustering*-based approach calculates dense or sparse optical flows and clusters them in order to identify moving object regions [17], [18].
- The *mosaiced background*-based approach creates a mosaiced background image and then uses a background subtraction technique to extract moving object regions [19]–[24].
- The *background compensation*-based approach estimates the transformation parameters between consecutive images by using corresponding features extracted from these images and creates a difference image in order to detect moving object regions [16], [25]–[27].

The latter approach is widely adopted, because it requires less computational cost and memory storage compared to the other two approaches. Most of these methods deal with PT cameras, and some of them are applicable to a PTZ camera. Indeed, in the case of zooming, the motion parallax problem arises, where the apparent motion of objects closer to the image plane is higher than that of objects that are further away. Instead, when there is no motion parallax, the apparent

A. Ferone is with the Department of Applied Science, University of Naples Parthenope, Centro Direzionale, Isola C/4, 80143 Naples, Italy. Phone +39 081 5476601, Fax +39 081 5476514, e-mail: alessio.ferone@uniparthenope.it.

L. Maddalena is with the Institute for High-Performance Computing and Networking, National Research Council, Via P. Castellino 111, 80131 Naples, Italy. Phone +39 081 6139522, Fax +39 081 6139531, e-mail: lucia.maddalena@cnr.it.

motion of all objects in the scene does not depend on their distance from the camera. This can be guaranteed by rotating the camera around its optical center (approach taken by many commercial systems) and holds also for most cameras when objects are far from it [19].

We afford the problem of moving object detection for PTZ cameras based on the adoption of artificial neural networks (ANNs), which are among the soft computing tools most frequently adopted for several video surveillance tasks, due to their well-known advantages, such as adaptivity and learning [28]. Indeed, an ANN can modify its connection weights using some training algorithms or learning rules; by updating the weights, the ANN can optimize its connections to adapt to changes in the environment. The capability of neural networks in emulating many unknown functional links by learning offline a limited set of representative examples allows one to infer a function from observations. This allows one to learn representations of the input that capture the salient input distribution features.

Neural network-based solutions to moving object detection have received considerable attention due to the fact that these methods are usually more effective and efficient than traditional ones [29]–[36]. Here we propose a neural-based background subtraction approach to moving object detection in image sequences taken from PTZ cameras, where the background model automatically adapts in a self-organizing way to the scene background variations. Such variations can be both those arising in a usual stationary camera setting (i.e., changes due to gradual illumination variations, to waving trees, or to shadows cast by moving objects [7]) and those due to the PTZ camera movement. The first type of variations are accurately handled by a neural background model for stationary cameras that has proven to accurately model image sequences and their variations in time [32]; this provides a background model particularly suitable for moving object detection, allowing us to robustly deal with typical problems of background subtraction. Handling of variations due to the camera movement is ensured by a novel registration mechanism applied to the neural background model in order to automatically compensate the eventual ego-motion, suitably estimated at each time instant. Therefore, contrary to previous background compensation-based approaches to moving object detection for PTZ cameras, background subtraction is achieved by an accurate and well-settled model, suitably adapted to the problem at hand, and not by frame differencing, which is notoriously sensitive to noise and variations in illumination and subject to the aperture problem [2], [5].

The paper is organized as follows. In Section II, we describe the neural self-organizing model for image sequences, and describe how such model is used for background modeling and how it is adapted to handle background variations due to the PTZ camera movement. In Section III, we present results achieved with the implementation of the proposed approach in terms of attained accuracy, comparing them with those obtained by several state-of-the-art methods. Section IV includes concluding remarks.

## II. Self-Organizing Background Subtraction for Video Sequences Taken by PTZ Cameras

The basic idea of our approach to moving object detection in image sequences taken from PTZ cameras consists of exploiting the available knowledge concerning the self-organized learning behavior of the brain, which is the foundation of human visual perception. Indeed, this brain behavior has been studied for a long time by many people, starting with [37], finding that topographically ordered maps are widely observed in the brain cortex. The main structures of the cortical maps are established before birth in a predetermined topographically ordered fashion; other more detailed areas (associative areas), however, are developed through self-organization gradually during life and in a topographically meaningful order [28]. Taking into account such topographically ordered projections is undoubtedly important for understanding and constructing dimension-reduction mappings and for the effective representation of sensory information and feature extraction. Therefore, relying on recent research in this area [32], we build the sequence background model by learning in a self-organizing manner image sequence variations, seen as trajectories of pixels in time. A neural network mapping method is proposed to use a whole trajectory incrementally in time fed as an input to the network. Each neuron computes a function of the weighted linear combination of incoming inputs, and therefore can be represented by a weight vector, obtained collecting the weights related to incoming links. An incoming pattern is mapped to the neuron whose set of weight vectors is most similar to the pattern, and weight vectors in a neighborhood of such node are updated. The obtained self-organizing neural network is organized as a 2-D grid of neurons, producing a representation of training samples with lower dimensionality, at the same time preserving topological neighborhood relations of the input patterns. Differently from [32], at each time instant the neural background model automatically compensates the eventual ego-motion due to the PTZ camera.

### A. Neural Model Representation

Given an image sequence $\{I_t\}$, for each pixel $\mathbf{p}$ in the image domain $D$, we build a neuronal map consisting of $n \times n$ weight vectors $m_t^{i,j}(\mathbf{p}), i, j = 0, \ldots, n-1$, which will be called a *model* for pixel $\mathbf{p}$ and will be indicated as $M_t(\mathbf{p})$:

$$M_t(\mathbf{p}) = \left\{ m_t^{i,j}(\mathbf{p}),\ i,j = 0, \ldots, n-1 \right\}. \tag{1}$$

If every sequence frame has $N$ rows and $P$ columns, the complete set of models $M_t(\mathbf{p})$ for all pixels $\mathbf{p}$ of the $t$th sequence frame $I_t$ is organized as a 2D neuronal map $B_t$ with $n \times N$ rows and $n \times P$ columns, where the weight vectors $m_t^{i,j}(\mathbf{p})$ for the generic pixel $\mathbf{p} = (x, y)$ are at neuronal map position $(n \times x + i, n \times y + j)$, $i, j = 0, \ldots, n-1$:

$$B_t(n \times x + i, n \times y + j) = m_t^{i,j}(\mathbf{p}),\ i,j = 0, \ldots, n-1. \tag{2}$$

This configuration of the whole neuronal map $B_t$ allows us to easily take into account the spatial relationship among pixels and corresponding weight vectors, as we shall see in the following subsections.

We explicitly observe that the notation introduced in (1) and (2) appears redundant. However, in the sequel we will adopt both notations:

- the model $M_t(\mathbf{p})$ for pixel $\mathbf{p}$ will be adopted to indicate the whole set of weight vectors for each single pixel, helping to focalize on the pixelwise representation of the background model;
- the neuronal map $B_t$ will be adopted to refer to the whole background model for an image sequence, to highlight spatial relationships among weight vectors of adjacent pixels, and to detail the implementation of the proposed moving object detection algorithm. Indeed, besides its role as a background model, the neuronal map $B_t$ can be thought of as a 2D array that stores, at each time $t$, all the weight vectors of the background model needed for background subtraction.

### B. Neural Model Initialization

In the case of our background modeling application, we have at our disposal a fairly good means of initializing the weight vectors of the network, because the first image of the sequence $I_0$ is indeed a good initial approximation of the background. Therefore, for each pixel $\mathbf{p}$, the corresponding weight vectors of the model $M_0(\mathbf{p})$ are initialized with the pixel brightness value at time $t = 0$:

$$m_0^{i,j}(\mathbf{p}) = I_0(\mathbf{p}), \quad i,j = 0,\ldots,n-1. \tag{3}$$

Therefore, the resulting neuronal map $B_0$, obtained for all pixels $\mathbf{p}$ as in (2) with $t$=0, using the values specified in (3), can be seen as an $n \times n$ enlarged version of the first sequence frame $I_0$. However, it is important to observe that the background model, arranged as a 2-D grid of 3-D weight vectors, could be seen as an image map at pixel super-resolution only at this stage. Indeed, as it will be clear in the next paragraph, the neural model update is carried out following the connections between neurons into the neuronal map.

### C. Background Subtraction and Neural Model Update

At each subsequent time step $t$, background subtraction is achieved by comparing each pixel of the $t$th sequence frame $I_t$ with the model for that pixel.

For the case of image sequences taken from stationary cameras [32], each incoming pixel $\mathbf{p}$ of $I_t$ is compared to the current pixel model $M_{t-1}(\mathbf{p})$ to determine if there exists a best matching weight vector $BM(\mathbf{p})$ that is close enough to it. If no acceptable matching weight vector exists, $\mathbf{p}$ is detected as belonging to a moving object (foreground). Otherwise, if such weight vector is found, it means that $\mathbf{p}$ is a background pixel. In the latter case, further learning of the neuronal map allows the background model to adapt to slight scene modifications. Such learning is achieved by updating the neural weights according to a visual attention mechanism of reinforcement, where the best matching weight vector, together with its neighborhood, is reinforced into the neuronal map. This step is not simply a modification of few neighboring pixels, but an adaptation of the weights, in a precise topological neighborhood, determined by connections between neurons into the neuronal map.

In the more general case of image sequences taken from PTZ cameras, the incoming pixel $\mathbf{p}$ of the $t$th sequence frame $I_t$ could have moved as compared to the previous time $t$-1. Therefore, the current model $M_{t-1}(\mathbf{p})$, whose weight vectors are stored in $B_{t-1}$ as described in (2), could be an improper model for actual pixel $\mathbf{p}$. In order to keep track of such spatial movements, we apply a novel registration mechanism to the background model. To this end, we compute the homography $H$ between sequence frames $I_{t-1}$ and $I_t$, that allows us to obtain, for each pixel $\mathbf{p}'$ of $I_{t-1}$, the corresponding pixel $\mathbf{p}$=$H\mathbf{p}'$ of $I_t$, and to exploit this information in order to address the proper model for current pixel $\mathbf{p}$.

The above described background subtraction and update procedure for each pixel can be sketched as the algorithm named PTZ-SOBS (Self-Organizing Background Subtraction for video sequences taken by PTZ cameras) reported in Fig. 1. Contrary to SOBS algorithm [32], we do not distinguish

**PTZ-SOBS algorithm**

Input: value $I_t(\mathbf{p})$ of pixel $\mathbf{p}$ in frame $I_t$, $t = 0, \ldots, T$.
Output: background model $M_t(\mathbf{p})$ for pixel $\mathbf{p}$ at time $t$;
detection mask value $D_t(\mathbf{p})$ for pixel $\mathbf{p}$ at time $t$.

1. *Initialize background model* $M_0(\mathbf{p})$ as in (3)
2. **for** $t = 1, T$
3. *Compute homography* $H$ *between* $I_{t-1}$ *and* $I_t$
4. **if** ($\exists \mathbf{p}'$ s.t. $\mathbf{p}$=$H\mathbf{p}'$) **then**
5. *Find in* $M_{t-1}(\mathbf{p}')$ *the best match* $BM(\mathbf{p})$ *to* $I_t(\mathbf{p})$
6. **if** ($BM(\mathbf{p})$ *close enough to* $I_t(\mathbf{p})$) **then**
7. $D_t(\mathbf{p})$ = 0 *//background*
8. $M_t(\mathbf{p}) = update(M_{t-1}(\mathbf{p}'), I_t(\mathbf{p}))$
9. **else**
10. $D_t(\mathbf{p})$ = 1 *//foreground*
11. $M_t(\mathbf{p})$ = $M_{t-1}(\mathbf{p}')$
12. **else**
13. *Initialize background model* $M_t(\mathbf{p})$

Fig. 1. PTZ-SOBS algorithm.

the whole process into the two *calibration* and *online* phases. Indeed, although the calibration phase, involving the neural network initial learning, could be beneficial in order to obtain a much richer initial background model, in the case of image sequences taken from PTZ cameras we cannot assume a sufficient number of initial frames that are free of moving foreground objects. All the steps of PTZ-SOBS algorithm not yet detailed in our previous description will be thoroughly analyzed in the following subsections.

*1) Computing the Homography H Between $I_{t-1}$ and $I_t$:* In the case of a PTZ camera, consecutive frames are related by a homography [38] that is a $3 \times 3$ matrix which maps points on a plane in one frame into points on a plane in the other frame.

Given a set of points $\mathbf{p}_i$ in $I_t$ and a corresponding set of points $\mathbf{p}'_i$ in $I_{t-1}$, the task is to compute the projective

transformation, a $3 \times 3$ matrix $H$, that maps each $\mathbf{p}'_i$ to $\mathbf{p}_i$, i.e., $\mathbf{p}_i = H\mathbf{p}'_i$, for each i.
In order to compute $H$ we employ a feature based approach using Scale Invariant Feature Transform (SIFT) keypoints [39]. SIFT keypoints are extracted from each new frame and then matched to those extracted from the previous frame. The homography $H$ is then estimated by means of Direct Linear Transformation [38] and Random Sample Consensus [40] algorithms that yield an initial guess for $H$ and a list of inlier matches. The initial estimated homography $H$ is further refined by using Levemberg-Marquardt optimization minimizing the reprojection error.

*2) Finding the Best Match $BM(\boldsymbol{p})$ to $I_t(\boldsymbol{p})$:* In the case that current pixel $\mathbf{p}$ in $I_t$ corresponds to a pixel $\mathbf{p}'$ in $I_{t-1}$, that is, $\mathbf{p} = H\mathbf{p}'$ (see line 4 of PTZ-SOBS algorithm), the value $I_t(\mathbf{p})$ is compared to the current pixel model, given by $M_{t-1}(\mathbf{p}')$, to determine the weight vector $BM(\mathbf{p})$ that best matches it:

$$d(BM(\mathbf{p}), I_t(\mathbf{p})) = \min_{i,j=0,\ldots,n-1} d(m^{i,j}_{t-1}(\mathbf{p}'), I_t(\mathbf{p})) \quad (4)$$

where the metric $d(\cdot,\cdot)$ is suitably chosen according to the specific color space being considered. Example metrics could be the Euclidean distance in RGB color space, or the Euclidean distance of vectors in the HSV color hexcone, as suggested in [41]. The latter is the one adopted for the experiments reported in Section III. Indeed, the HSV color space allows one to specify colors in a way that is close to human experience of colors, relying on the hue, saturation, and value properties of each color. Moreover, hue stability against illumination changes is known to be important both for cast shadow suppression [42] and for motion analysis [43]–[45].

The best matching weight vector $BM(\mathbf{p})$, computed as in (4), is considered *close enough* to the pixel value $I_t(\mathbf{p})$ (see line 6 of PTZ-SOBS algorithm) if

$$d(BM(\mathbf{p}), I_t(\mathbf{p})) \leq \epsilon \quad (5)$$

where $\epsilon$ is a threshold allowing one to distinguish between foreground and background pixels. High values for $\epsilon$ allow us to obtain a background model including several observed pixel intensity variations, but these could be also due to moving objects. On the other side, lower values allow us to obtain a background model with less spurious pixels due to moving objects, at the price of lower detection accuracy.

If $BM(\mathbf{p})$ satisfies (5) then $\mathbf{p}$ is detected as a background pixel (see line 7 of PTZ-SOBS algorithm); otherwise it is detected as a foreground pixel (see line 10 of PTZ-SOBS algorithm).

*3) Updating the Model for pixel $\boldsymbol{p}$:* If $\mathbf{p}$ is detected as a background pixel, then its model should be updated, while if it is a foreground pixel, no update should be applied. This selectivity allows the background model to adapt to scene modifications, such as gradual light changes, without introducing the contribution of pixels that do not belong to the background scene, such as moving or standing persons and abandoned objects [46]. However, in both cases modifications should be applied to the model of $\mathbf{p}$ in order to take into account the eventual movement of the PTZ camera from the previous frame.

If an acceptable best matching weight vector $BM(\mathbf{p})$ is found for current sample $\mathbf{p}$, satisfying (5), the current pixel model, given by $M_{t-1}(\mathbf{p}')$, should be updated (see line 8 of PTZ-algorithm). To this end, the weight vectors of $B_{t-1}$ in a neighborhood of $BM(\mathbf{p})$ are updated according to selective weighted running average. In details, if $BM(\mathbf{p})$ is found at position $\overline{\mathbf{p}}$ in $B_{t-1}$, then weight vectors of $B_{t-1}$ are updated according to

$$B_t(\mathbf{q}) = (1-\alpha(\overline{\mathbf{p}}, \mathbf{q}))B_{t-1}(\mathbf{q}) + \alpha(\overline{\mathbf{p}}, \mathbf{q})I_t(\mathbf{p}) \quad \forall \mathbf{q} \in N_{\overline{\mathbf{p}}}. \quad (6)$$

Here $N_{\overline{\mathbf{p}}} = \{\mathbf{q} : |\overline{\mathbf{p}} - \mathbf{q}| < w\}$ is a 2D spatial neighborhood of $\overline{\mathbf{p}}$ of size $(2w-1)\times(2w-1)$ including $\overline{\mathbf{p}}$. Moreover, $\alpha(\overline{\mathbf{p}}, \mathbf{q}) = \gamma \cdot G(\mathbf{q} - \overline{\mathbf{p}})$, where $\gamma$ represents the learning rate, that depends on the scene variability, while $G(\cdot) = \mathcal{N}(\cdot; \mathbf{0}, \sigma^2 I)$ is a 2D Gaussian low-pass filter [47] with zero mean and $\sigma^2 I$ variance. The $\alpha(\overline{\mathbf{p}}, \mathbf{q})$ values are weights that allow us to smoothly take into account the spatial relationship between current pixel $\mathbf{p}$ (through its best matching weight vector found at position $\overline{\mathbf{p}}$) and its neighboring pixels in $I_t$ (through weight vectors $\mathbf{q} \in N_{\overline{\mathbf{p}}}$), thus preserving topological properties of the input (close inputs correspond to close outputs). Large $\gamma$ values enable the network to faster learn changes corresponding to background, but leading to faster inclusion into the background model of pixels belonging to foreground moving objects that have erroneously been detected as background (false negatives). On the other hand, lower learning rates make the network slower to adapt to rapid background changes, but making the model more tolerant to errors due to false negatives. Indeed in this case the problem is more easily corrected through self-organization, since weight vectors of false negative pixels are readily smoothed out by the learning process itself. In summary, the update in (6) allows us to include into the neural model the temporal variations of the scene background, at the same time taking into account, for each pixel, the variations detected in adjacent pixels.

If an acceptable best matching weight vector $BM(\mathbf{p})$ is not found for current sample $\mathbf{p}$, the background model $B_{t-1}$ should not be updated. In order to take into account the eventual movement of the PTZ camera from the previous frame, the model $M_{t-1}(\mathbf{p}')$ is just copied to the actual model $M_t(\mathbf{p})$ (see line 11 of PTZ-SOBS algorithm). Using (2) and supposing $\mathbf{p} = H\mathbf{p}'$, with $\mathbf{p} = (x, y)$ and $\mathbf{p}' = (x', y')$, the neuronal map $B_t$ is therefore updated by:

$$B_t(n \times x + i, n \times y + j) = B_{t-1}(n \times x' + i, n \times y' + j)$$

for $i, j = 0, \ldots, n-1$.

*4) Initializing the Model $M_t(\boldsymbol{p})$ for New Background Pixels:* In the case that the current pixel $\mathbf{p}$ in $I_t$ is a new background pixel, and therefore it does not correspond to any pixel $\mathbf{p}'$ in $I_{t-1}$ (see line 12 of PTZ-SOBS algorithm), the background model $M_t(\mathbf{p})$ should be somehow initialized. Such an initialization is usually left unspecified in most literature on background compensation. However, its role for moving object detection is very important, in order to not introduce non-existent moving objects into new background areas.

A simple initialization as the one proposed in (3) for $M_0(\mathbf{p})$ could introduce moving objects into the background

model. Instead, we propose to combine the actual frame pixel value $I_t(\mathbf{p})$ with the previous value of the background model $M_{t-1}(\mathbf{p})$ for the same pixel $\mathbf{p} = (x, y)$, by setting:

$$B_t(n \times x + i, n \times y + j) = \\ = (1 - \beta)B_{t-1}(n \times x + i, n \times y + j) + \beta I_t(x, y) \quad (7)$$

for $i, j = 0, \ldots, n-1$, where $\beta$ is a suitable weighting factor.

### D. Other Key Issues

The description of the neural image sequence model and of its use for background modeling given in the previous subsections has taken into account only basic key issues concerning the background subtraction problem. For the sake of simplicity, we intentionally omitted the treatment of other aspects that still are important for real-world applications. Here we give hints on some of such supplementary key issues and on their handling using our proposed model.

*1) Cast Shadows:* One of the well known issues in background maintenance is that of cast shadows: the background model should include the shadow, cast by moving objects, that apparently behaves itself moving, in order to have a more accurate detection of the moving objects shape. As in [32], the approach here adopted for moving cast shadow detection is adapted from that reported in [42], that proved to be quite accurate and suitable for moving object detection. Once detected, cast shadow pixels are treated as background pixels, and the corresponding models are updated as specified by (6). Further details about handling cast shadows can be found in [32].

*2) Robustness Against False Detections:* In order to enhance robustness against false detections, as in [48] we exploit the notion of spatial coherence into the background subtraction process. To this end, we compute a Neighborhood Coherence Factor [49], that gives a relative measure of the number of pixels, belonging to the spatial neighborhood of a given pixel, whose value is well represented by the background model. Such factor is used for both foreground segmentation and in the background updating formula. It has been shown on publicly available datasets that the introduction of spatial coherence allows us to greatly enhance robustness of the background subtraction algorithm against false detections [46].

*3) Sudden Illumination Changes:* The SOBS algorithm was not originally designed to cope with sudden illumination changes. Indeed, its selective update of the background prevents the insertion into the background model of anything that differs too much from the background itself, including sudden illumination changes. Therefore, here we introduced an intensity stabilization technique for the background model. It consists in computing, at each time $t$, the average intensity difference $d_t$ in subsequent images $I_{t-1}$ and $I_t$, giving a rough estimation of the eventual global illumination change; if the estimated value of $d_t$ exceeds the foreground segmentation threshold, then $d_t$ is added to one of the weight vectors of the background model $B_{t-1}$ for each pixel. This ensures that at least one of the weight vectors takes into account the estimated global illumination change.

## III. Experimental Results

Several experiments have been conducted to validate our approach to moving object detection in image sequences taken from PTZ cameras and to compare its results with those achieved by other state-of-the-art methods. In the following, choice of parameter values, accuracy metrics, qualitative and quantitative results will be described for several publicly available sequences.

### A. Parameter Values

Values of PTZ-SOBS algorithm parameters experimentally chosen for all the considered sequences are reported in Table I. The distance threshold $\epsilon$ adopted in (5) strictly depends on the color similarity between the moving objects and the background; suitable values are in the order of $10^{-2}$. The learning rate $\gamma$ for the background update in (6) depends on the scene variability and, therefore, also on the PTZ movement speed. Indeed, in the case of moving cameras, values higher than for static cameras (ranging in [$10^{-2}$, $10^{-1}$], as reported in [32]) are needed in order to achieve high accuracy. The weighting factor $\beta$ for initializing the background model of new background pixels in (7) depends on the content of the newly framed parts of the scene. Indeed, if the new scene parts contain previously unseen moving objects, then values of $\beta$ lower than 1.0 limit their inclusion into the corresponding new background model.

TABLE I
PARAMETER VALUES FOR PTZ-SOBS ALGORITHM FOR THE CONSIDERED IMAGE SEQUENCES.

| | $\epsilon$ | $\gamma$ | $\beta$ |
|---|---|---|---|
| *Backyard* | 0.070 | 0.60 | 0.55 |
| *People1* | 0.008 | 0.36 | 0.95 |
| *People2* | 0.013 | 0.45 | 0.70 |
| *Cars1* | 0.008 | 0.24 | 0.94 |
| *Cars6* | 0.011 | 0.47 | 0.80 |

The remaining parameters have been fixed for all the sequences as $n = 3$ ($n^2$ being the number of weight vectors used to model each pixel), $w = 1$, and $\sigma = 0.75$ ($w$ and $\sigma^2$ being the half-width of the neighborhood $N_{\mathbf{p}}$ and the variance $\sigma^2$ of the 2D Gaussian needed to specify the weights for the running average in (6), respectively). The choice has been driven by experiments carried out, where, varying the parameters, we observed almost constant accuracy.

### B. Accuracy Metrics

The adopted metrics for evaluating the accuracy are *Recall* (Rec), *Precision* (Prec), and *F-measure* ($F_1$), defined as

$$\mathrm{Rec} = \frac{TP}{TP + FN}, \quad \mathrm{Prec} = \frac{TP}{TP + FP}$$

and

$$F_1 = \frac{2 \times \mathrm{Rec} \times \mathrm{Prec}}{\mathrm{Rec} + \mathrm{Prec}}$$

where *TP, FN*, *FP* are the total number of true positive, false negative, and false positive pixels, respectively. Such values have been obtained comparing the ground truth masks with the corresponding computed segmentation results.


Fig. 2. Results of moving object detection for frames 489, 507, 595, 748, and 768 of the *Backyard* sequence: (a) original frames; (b) ground truth masks; (c) moving object detection masks computed by the PTZ-SOBS algorithm; (d) representation of the neural background model.

### C. Results on MICC Dataset

The *Backyard* sequence belongs to a dataset for PTZ camera pose recovery[1] of the Media Integration and Communication Center (MICC) [50]. It is an outdoor sequence consisting of 831 frames of 320 $\times$ 240 spatial resolution, adopted here for showing the results achieved by the PTZ-SOBS algorithm under difficult illumination and cluttered background conditions. The scene consists of people walking down a street, with parked cars and waving trees, taken by a moving camera.

In Fig. 2 we report the original sequence frames no. 489, 507, 595, 748, and 768 (column (a)), the corresponding ground truth masks[2] (column (b)), the moving object detection mask computed by the PTZ-SOBS algorithm (column (c)), and a representation of the compensated background model (column (d)). In frame 489 (the fifth frame considered for our experiments) there are no moving people, and the camera has slightly panned from left to right; the compensated background model provides quite a good representation of the real background, and the corresponding detection mask is correctly empty. In frame 507 a person has entered the scene and a moving tree appears in the top left of the scene. The compensated background model provides a good representation of the real background, also demonstrating the robustness of the adopted neural model to moving background, and allows us to obtain a quite accurate detection mask. In frame 595 another person entered the scene and is currently under a wide branch of a waving tree dominating the top left part of the scene. The background model still allows us to obtain a quite accurate moving object detection mask, despite the presence of the waving tree. In frame 748 the scene view is completely changed due to the PTZ camera movement, and again the background model and the corresponding moving object detection mask appear quite accurate. Finally, in frame 768 the moving person is walking down the street in an area where illumination conditions worsen, exhibiting a foggy area around the person head. Nonetheless, the background model still allows us to detect most part of the person body, even though it is barely visible by the human eye. It should be explicitly observed that all false negative pixels are due to evident camouflage of the scene with some moving person details (the scarf with the pavement in frame 507, the trouser with the moving tree in frame 595, and the befogged face with the pavement in frame 768), and could be avoided by connected component analysis of the detection mask, here intentionally omitted.

Average segmentation accuracy results achieved by the proposed approach on the *Backyard* sequence are quite high (Prec = 0.984, Rec = 0.660, $F_1$ = 0.788). The achieved high *Recall* value ensures that most of the moving pixels are detected as moving; at the same time, the high *Precision* value indicates that only few of the pixels detected as moving are

[1]Publicly available at http://www.micc.unifi.it/vim/datasets/ptz-camera-pose-recovery/, last accessed on 09/10/12.

[2]We made publicly available the hand-segmented ground truth masks in the download section of http://cvprlab.uniparthenope.it.

instead background pixels. The consequent high values of $F_1$, that is the weighted harmonic mean of *Precision* and *Recall*, allow us to deduce the overall high segmentation accuracy of the proposed approach for this sequence.

### D. Results on Hopkins 155 Dataset

Some of the sequences from the Berkeley Motion Segmentation Dataset[3] [11], taken from the Hopkins 155 dataset [51], have been considered in order to compare the results achieved by the proposed method with several state-of-the-art methods.

The chosen sequences, named *People1, People2, Cars1* and *Cars6*, have been taken from hand-held cameras, but they exhibit a camera motion that can be assimilated to that of a PTZ camera. None of these sequences has initial empty frames, nor they have initial static frames that enable background estimation. Therefore, we provided an hand-made initial background estimate; other initialization strategies could have been adopted, such as in [13].

Results obtained on selected frames of the considered sequences, using the parameter values specified in Section III-A, are reported in Fig. 3, where we can observe the PTZ-SOBS detection accuracy. Indeed, the background model (shown in column (d)) is well adapted to the camera movement, and the detection masks (shown in column (c)) are quite similar to the provided ground truth masks (shown in column (b)).

In Table II we report average accuracy values achieved by the proposed approach on the sequences of Fig. 3, compared with those achieved by several state-of-the-art methods:

- In [9] Sheikh et al. use orthographic motion segmentation over a sliding window to segment a set of trajectories, followed by sparse modeling of per frame appearance to densely segment images. Their results in Table II have been reported in [13].
- The motion segmentation algorithm developed by Brox and Malik [11] analyzes the point trajectories along the sequence and segments them into clusters. To obtain pixel-level segmentation, and the corresponding results reported in Table II, the variational method [52] has been applied by Zhou et al. in [14] to turn the trajectory clusters into dense regions.
- In the work of Kwack et al. [12] foreground and background appearance models in each frame are constructed and propagated sequentially by Bayesian filtering. The motion, which transfers the previous appearance models to the current frame, is robustly estimated by nonparametric belief propagation (NBP). Their results, both including and excluding NBP, have been reported in [12].
- Elqursh and Elgammal [13] present a method which enables learning of pixel-based foreground and background models and segments each frame in an online framework. Indeed, rather than producing a single segmentation as an output, it uses Bayesian filtering to maintain a belief over different labellings, and appearances of the background and foreground, thus allowing the recovering from errors. Their results have been reported in [14] for the proposed approach with and without the label prior (indicated in Table II as method 1 and 2, respectively).
- In order to segment moving objects from image sequences, Zhou et al. [14] avoid complicated motion computation by formulating the problem as outlier detection, integrating object detection and background learning into a single process of optimization. Their results in Table II have been reported in [14].

From Table II we can observe that PTZ-SOBS achieves the highest F-measure values as compared to all the considered methods for the sequences *People1, People2*, and *Cars6*, while for sequence *Cars1*, although quite high, they are lower than those achieved by [12] and [13].

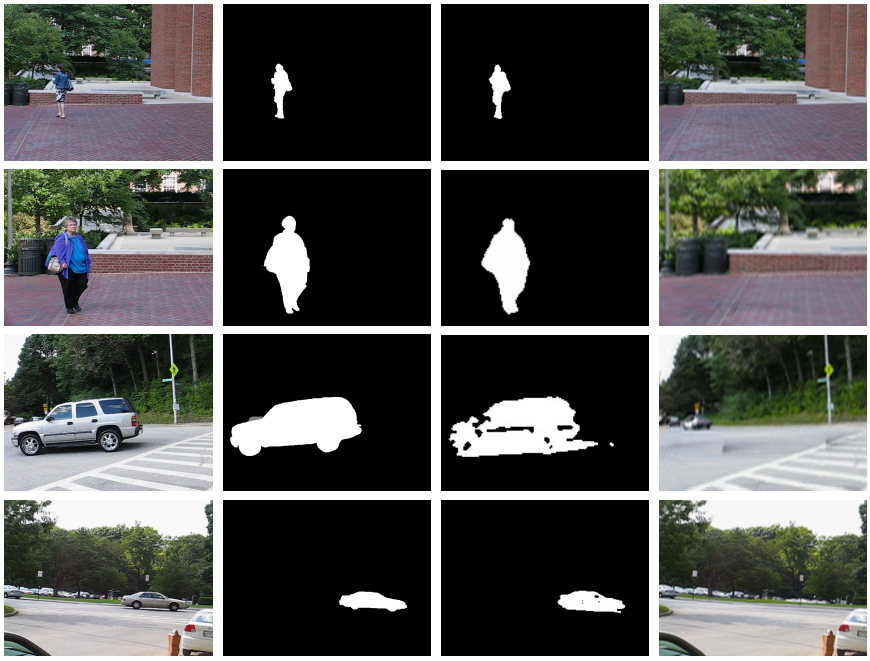
TABLE II
COMPARISON OF AVERAGE PRECISION, RECALL, AND F-MEASURE VALUES ON THE SEQUENCES OF FIG. 3.

| | *People1* | | | *People2* | | |
|---|---|---|---|---|---|---|
| | Prec | Rec | $F_1$ | Prec | Rec | $F_1$ |
| PTZ-SOBS | 0.958 | 0.923 | 0.940 | 0.931 | 0.971 | 0.950 |
| Sheikh et al. [9] | 0.780 | 0.630 | 0.697 | 0.730 | 0.830 | 0.777 |
| Brox and Malik [11] | 0.890 | 0.775 | 0.829 | 0.917 | 0.892 | 0.904 |
| Kwak et al. [12] - with NBP | 0.950 | 0.930 | 0.940 | 0.850 | 0.890 | 0.870 |
| Kwak et al. [12] - without NBP | 0.910 | 0.760 | 0.828 | 0.910 | 0.920 | 0.915 |
| Elqursh and Elgammal [13] - 1 | 0.940 | 0.850 | 0.893 | 0.690 | 0.880 | 0.774 |
| Elqursh and Elgammal [13] - 2 | 0.970 | 0.880 | 0.923 | 0.870 | 0.880 | 0.875 |
| Zhou et al. [14] | 0.936 | 0.933 | 0.934 | 0.925 | 0.965 | 0.945 |
| | *Cars1* | | | *Cars6* | | |
| | Prec | Rec | $F_1$ | Prec | Rec | $F_1$ |
| PTZ-SOBS | 0.806 | 0.927 | 0.862 | 0.866 | 0.964 | 0.913 |
| Sheikh et al. [9] | 0.630 | 0.990 | 0.770 | N.A. | N.A. | N.A. |
| Brox and Malik [11] | N.A. | N.A. | N.A. | 0.824 | 0.994 | 0.901 |
| Kwak et al. [12] - with NBP | 0.920 | 0.840 | 0.878 | N.A. | N.A. | N.A. |
| Kwak et al. [12] - without NBP | 0.410 | 0.220 | 0.286 | N.A. | N.A. | N.A. |
| Elqursh and Elgammal [13] - 1 | 0.840 | 0.990 | 0.909 | N.A. | N.A. | N.A. |
| Elqursh and Elgammal [13] - 2 | 0.850 | 0.970 | 0.906 | N.A. | N.A. | N.A. |
| Zhou et al. [14] | N.A. | N.A. | N.A. | 0.837 | 0.984 | 0.905 |

## IV. CONCLUSIONS

We propose an approach to the problem of moving object detection in image sequences taken from PTZ cameras based on the idea of exploiting the available knowledge concerning the self-organized learning behavior of the brain, which is the foundation of human visual perception, and traducing it into models and algorithms that can accurately solve the problem. An extension to the case of PTZ cameras of a neural self-organizing background model is proposed, that automatically adapts to variations in the scene background, both arising in a usual stationary camera setting, and due to the PTZ camera movement. The compensated background model is adopted for accurate moving object detection, as demonstrated by experimental results on real image sequences.

[3]Publicly available at http://lmb.informatik.uni-freiburg.de/resources/datasets/, together with binary ground truth images, last accessed on 09/10/12.



(a) (b) (c) (d)

Fig. 3. Results of moving object detection for frame 40 of the *People1* sequence (first row), frame 30 of the *People2* sequence (second row), frame 10 of the *Cars1* sequence (third row), frame 30 of the *Cars6* sequence (fourth row): (a) original frames; (b) ground truth masks; (c) moving object detection masks computed by the PTZ-SOBS algorithm; (d) representation of the neural background model.

## REFERENCES


[1] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, and L. Wixson, "A system for video surveillance and monitoring," Tech. Rep. CMU-RI-TR-00-1212, The Robotics Inst., Carnegie Mellon Univ., Pittsburgh, PA, 2000.

[2] S.-C. S. Cheung and C. Kamath, "Robust techniques for background subtraction in urban traffic video," in *Proc. EI-VCIP*, S. Panchanathan and B. Vasudev, Eds., vol. 5308. SPIE, 2004, pp. 881–892.

[3] M. Piccardi, "Background subtraction techniques: A review," in *Proc. IEEE Int. Conf. Systems, Man, Cybernetics*, 2004, pp. 3099–3104.

[4] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 294–307, May 2005.

[5] S. Elhabian, K. El Sayed, and S. Ahmed, "Moving object detection in spatial domain using background removal techniques: State-of-art," *Recent Patents on Computer Science*, vol. 1, no. 1, pp. 32–54, Jan. 2008.

[6] S. Brutzer, B. Hoferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *Proc. CVPR*, 2011, pp. 1937–1944.

[7] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: principles and practice of background maintenance," in *Proc. ICCV*, 1999, pp. 255–261.

[8] K. Huang, L. Wang, T. Tan, and S. Maybank, "A real-time object detecting and tracking system for outdoor night surveillance," *Pattern Recogn.*, vol. 41, no. 1, pp. 432–444, 2008.

[9] Y. Sheikh, O. Javed, and T. Kanade, "Background subtraction for freely moving cameras," in *Proc. ICCV*, 2009, pp. 1219–1225.

[10] C. Xu, J. Liu, and B. Kuipers, "Motion segmentation by learning homography matrices from motor signals," in *Proc. Canadian Conf. on Computer and Robot Vision*, ser. CRV '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 316–323.

[11] T. Brox and J. Malik, "Object segmentation by long term analysis of point trajectories," in *Proc. ECCV*, 2010, pp. 282–295.

[12] S. Kwak, T. Lim, W. Nam, B. Han, and J. H. Han, "Generalized background subtraction based on hybrid inference by belief propagation and bayesian filtering," in *Proc. ICCV*, 2011, pp. 2174–2181.

[13] A. Elqursh and A. Elgammal, "Online moving camera background subtraction," in *Computer Vision - ECCV 2012*, ser. Lecture Notes in Computer Science, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Springer Berlin / Heidelberg, 2012, vol. 7577, pp. 228–241.

[14] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PP, no. 99, p. 1, 2012.

[15] C. Micheloni, B. Rinner, and G. Foresti, "Video analysis in pan-tilt-zoom camera networks," *IEEE Signal Process. Mag.*, vol. 27, no. 5, pp. 78–90, Sept. 2010.

[16] J. K. Suhr, H. G. Jung, G. Li, S.-I. Noh, and J. Kim, "Background compensation for pan-tilt-zoom cameras using 1-d feature matching and outlier rejection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 3, pp. 371–377, March 2011.

[17] Y.-K. Jung, K.-W. Lee, and Y.-S. Ho, "Feature-based object tracking with an active camera," in *Proc. IEEE Pacific Rim Conf. on Multimedia: Advances in Multimedia Information Processing*, ser. PCM '02. London, UK, UK: Springer-Verlag, 2002, pp. 1137–1144.

[18] P. Varcheie and G.-A. Bilodeau, "Adaptive fuzzy particle filter tracker for a PTZ camera in an IP surveillance system," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 2, pp. 354–371, Feb. 2011.

[19] A. Mittal and D. Huttenlocher, "Scene modeling for wide area surveillance and image synthesis," in *Proc. CVPR*, vol. 2, 2000, pp. 160–167.

[20] E. Hayman and J.-O. Eklundh, "Statistical background subtraction for a mobile observer," in *Proc. ICCV*, vol. 1, 2003, pp. 67–74.

[21] Y. Ren, C.-S. Chua, and Y.-K. Ho, "Statistical background modeling for non-stationary camera," *Pattern Recognition Letters*, vol. 24, no. 1-3, pp. 183–196, 2003.

[22] A. Bartoli, N. Dalal, and R. Horaud, "Motion panoramas," *Computer Animation and Virtual Worlds*, vol. 15, no. 5, pp. 501–517, 2004.

[23] A. Bevilacqua and P. Azzari, "High-quality real time motion detection using PTZ cameras," in *IEEE Int. Conf. on Video and Signal Based Surveillance*, Nov. 2006, p. 23.

[24] ——, "A fast and reliable image mosaicing technique with application to wide area motion detection," in *Image Analysis and Recognition*, ser. Lecture Notes in Computer Science, M. Kamel and A. Campilho, Eds. Springer Berlin Heidelberg, 2007, vol. 4633, pp. 501–512.

[25] S. Araki, T. Matsuoka, H. Takemura, and N. Yokoya, "Real-time tracking of multiple moving objects in moving camera image sequences using robust statistics," in *Proc. ICPR*, vol. 2, Aug. 1998, pp. 1433–1435.

[26] C. Micheloni and G. L. Foresti, "Real-time image processing for active monitoring of wide areas," *J. of Visual Communication and Image Representation*, vol. 17, no. 3, pp. 589–604, 2006.

[27] X. D. Pham, J. U. Cho, and J. W. Jeon, "Background compensation using hough transformation," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, May 2008, pp. 2392–2397.
[28] L. Maddalena and A. Petrosino, "Neural networks in video surveillance: A perspective view," in *Handbook on Soft Computing for Video Surveillance*, S. Pal, A. Petrosino, and L. Maddalena, Eds. Chapman & Hall/CRC, 2012, pp. 59–78.
[29] A. J. Schofield, P. A. Mehta, and T. J. Stonham, "A system for counting people in video images using neural networks to identify the background scene," *Pattern Recogn.*, vol. 29, no. 8, pp. 1421–1428, 1996.
[30] G. Pajares, "A hopfield neural network for image change detection," *IEEE Trans. Neural Netw.*, vol. 17, no. 5, pp. 1250–1264, sept. 2006.
[31] D. Culibrk, O. Marques, D. Socek, H. Kalva, and B. Furht, "Neural network approach to background modeling for video object segmentation," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1614–1627, nov. 2007.
[32] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1168–1177, Jul. 2008.
[33] M. Chacon, G. Sergio, and V. Javier, "Simplified SOM-neural model for video segmentation of moving objects," in *Int. Joint Conf. on Neural Networks*, June 2009, pp. 474–480.
[34] R. M. Luque, J. Ortiz-De-Lazcano-Lobato, E. Lopez-Rubio, and E. J. Palomo, "Object tracking in video sequences by unsupervised learning," in *Proc. Int. Conf. on Computer Analysis of Images and Patterns*, ser. CAIP '09. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 1070–1077.
[35] W. Zhiming, Z. Li, and B. Hong, "Pnn based motion detection with adaptive learning rate," in *Int. Conf. on Computational Intelligence and Security*, vol. 1, Dec. 2009, pp. 301–306.
[36] E. Lopez-Rubio, R. M. L. Baena, and E. Domnguez, "Foreground detection in video sequences with probabilistic self-organizing maps." *Int. J. Neural Syst.*, vol. 21, no. 3, pp. 225–246, 2011.
[37] D. Hebb, *The organization of behavior*. New York: Wiley & Sons, 1949.
[38] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.
[39] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[40] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.
[41] R. Fisher, "Change detection in color images," 1999. [Online]. Available: http://homepages.inf.ed.ac.uk/rbf/PAPERS/iccv99.pdf
[42] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1337–1342, Oct. 2003.
[43] J. Barron and R. Klette, "Experience with optical flow in colour video image sequences," in *Proc. IVC*, 2001, pp. 195–200.
[44] ——, "Quantitative color optical flow," in *Proc. ICPR*, vol. 4, 2002, pp. 251–255.
[45] Y. Mileva, A. Bruhn, and J. Weickert, "Illumination-robust variational optical flow with photometric invariants," in *Proc. 29th DAGM Conf. on Pattern Recognition*. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 152–162.
[46] L. Maddalena and A. Petrosino, "The SOBS algorithm: What are the limits?" in *Proc. CVPR Workshops*, june 2012, pp. 21–26.
[47] P. Burt, "Fast filter transform for image processing," *Computer Graphics and Image Processing*, vol. 16, no. 1, pp. 20–51, 1981.
[48] L. Maddalena and A. Petrosino, "A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection," *Neural Comput. Appl.*, vol. 19, no. 2, pp. 179–186, Mar. 2010.
[49] J. Ding, R. Ma, and S. Chen, "A scale-based connected coherence tree algorithm for image segmentation," *IEEE Trans. Image Process.*, vol. 17, no. 2, pp. 204–216, 2008.
[50] A. Del Bimbo, G. Lisanti, I. Masi, and F. Pernici, "Continuous recovery for real time pan tilt zoom localization and mapping," in *Proc. AVSS*, Sept. 2011, pp. 160–165.
[51] R. Tron and R. Vidal, "A benchmark for the comparison of 3-D motion segmentation algorithms," in *Proc. CVPR*, 2007, pp. 282–295.
[52] P. Ochs and T. Brox, "Object segmentation in video: A hierarchical variational approach for turning point trajectories into dense regions," in *Proc. ICCV*, 2011, pp. 1583–1590.

**Alessio Ferone** received the Laurea degree (cum laude) in computer science from the University of Naples Parthenope and the Ph.D in computer science from the University of Milan, Milan, Italy.

He is currently an assistant professor in computer science at the University of Naples Parthenope, Naples, Italy. His research interests include image processing, pattern recognition, neural networks, and multimedia systems.

Dr. Alessio Ferone serves as a reviewer for several international journals and is a Member of IAPR.

**Lucia Maddalena** (M'08) received the Laurea degree (cum laude) in mathematics and the Ph.D in applied mathematics and computer science from the University of Naples Federico II, Naples, Italy.

She is a researcher at the Institute for High-Performance Computing and Networking, National Research Council of Italy. Initial research dealt with parallel computing algorithms, methodologies, and techniques, and their application to computer graphics. Subsequent research is devoted to methods, algorithms, and software for image processing and multimedia systems in high-performance computational environments, with application to real-world problems, mainly digital film restoration and video surveillance.

Dr. Lucia Maddalena taught at the University of Naples Federico II and at the University of Naples Parthenope. She is an associate editor of the International Journal of Biomedical Data Mining, serves as a reviewer for several international journals, and is a Member of IAPR.

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF FOOTNOTES

1. Publicly available at http://www.micc.unifi.it/vim/datasets/ptz-camera-pose-recovery/, last accessed on 09/10/12. . . . . 6
2. We made publicly available the hand-segmented ground truth masks in the download section of http://cvprlab.uniparthenope.it. . . . . 6
3. Publicly available at http://lmb.informatik.uni-freiburg.de/resources/datasets/, together with binary ground truth images, last accessed on 09/10/12. . . 7