IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 14, 2017, accepted April 25, 2017, date of publication April 28, 2017, date of current version June 7, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2699227

Illumination-Invariant Background Subtraction : Comparative Review, Models, and Prospects

WONJUN KIM[1], (Member, IEEE), AND CHANHO JUNG[2]
[1]Department of Electronics Engineering, Konkuk University, Seoul 05029, South Korea
[2]Department of Electrical Engineering, Hanbat National University, Daejeon 34158, South Korea

Corresponding author: Chanho Jung (peterjung@hanbat.ac.kr)

ABSTRACT Background subtraction is a key prerequisite for a wide range of image processing applications due to its pervasiveness in various contexts. In particular, video surveillance highly requires the reliable background subtraction for further operations, such as object tracking and recognition, and thus, enormous efforts for this task have been devoted in recent decades. However, the path of technological evolution for background subtraction has now faced with an important issue that has started to be resolved: sensitivity to dynamic changes of scene contexts (e.g., illumination variations and moving backgrounds). Such dynamic changes are hardly tolerated by most of traditional background models, since they yield the drastically different statistics of pixel values even onto the relevant position between consecutive frames. To resolve this problem, many researchers in this field have developed robust and efficient methods. The goal of this paper is to provide a comprehensive review with a special attention to schemes related to handling varying illuminations frequently occurring in the outdoor surveillance scenario. This paper covers a systematic taxonomy, methodologies, and performance evaluations on benchmark databases, and also provides constructive discussions for the smart video surveillance under unconstrained outdoor environments.

INDEX TERMS Background subtraction, dynamic changes of scene contexts, varying illuminations, outdoor surveillance.

I. INTRODUCTION

With rapidly increasing the number of various cameras available on diverse environments, video surveillance has been actively studied in last decades. Furthermore, for the public safety against criminal attacks and terrorisms, more advanced operations such as object tracking, recognition, and behavior understanding have started to be applied to the surveillance system. For example, patterns of motions occurring in the particular region are accordingly analyzed to detect doubtful bags, abnormal events, etc. Background subtraction is an essential task for the success of video surveillance under a wide range of real-world scenarios. The purpose of background subtraction is to discriminate moving objects, which are referred to as foreground, from static parts of a given scene [1]. As a pioneer work, the difference between the current frame and the simple background model (e.g., single Gaussian model [2]) is computed at each pixel position to detect notable motions and the background model is subsequently updated to allow small variations of pixel values on the relevant position into the background model. While this static background model performs appropriately under constrained indoor environments, it yields significant problems in real-world situations including background clutters, outdoor dynamics, etc. Even though a substantial amount of algorithms has been proposed for background subtraction, most of them still need to be improved to cope with these limitations.

Recently, background subtraction for outdoor surveillance has gained great attentions due to demands on crowd analysis [3], abnormal event monitoring [4], intelligent traffic control systems [5], people safety [6], animal surveillance in the wild [7], etc. It is naturally considered a fact that the deployment of outdoor surveillance systems keeps growing year after year in such diverse environments as downtown, campus or border areas. Therefore, considerable efforts have been thrown into challenging tasks of background subtraction under outdoor environments, not only by academic fields, but also by private companies. The main problems to achieve the reliable outdoor surveillance are dynamic changes of background in a given scene, which can be briefly summarized as follows:

- *Change of lighting conditions* : while the illumination gradually changes according to the time in a day (i.e., the sun moves across the sky), it also can dynamically

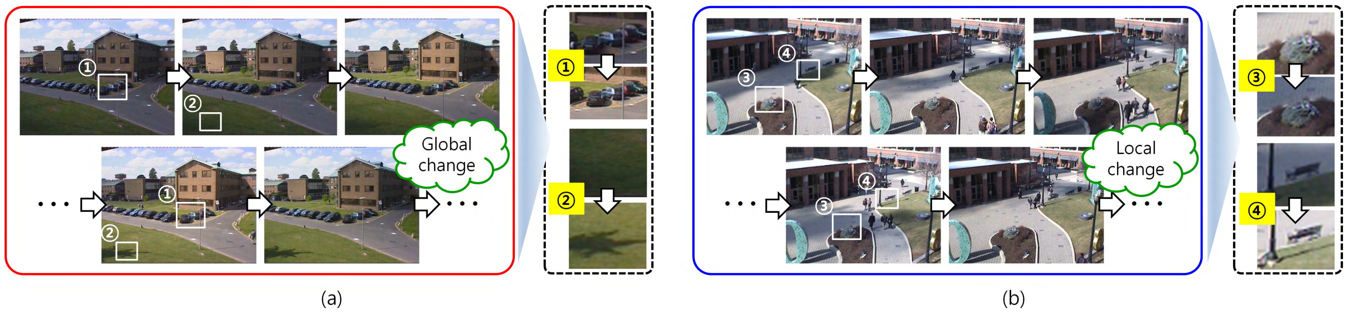


FIGURE 1. **Varying illuminations in outdoor environments. Some examples from (a) PETS2001 [82] and (b) OTCBVS [83] databases. Notice that the lighting condition changes globally in PETS2001 samples while it locally varies due to moving clouds in a short time. It should be emphasized that the dramatic changes occur even in the same background region by varying illuminations (see ①, ②, ③, and ④).**

varies by the movement of clouds in local as well as global areas.

- *Moving backgrounds* : structured and repeated motion patterns, e.g., waving trees and rippling water, often occur in outdoor environments, which are distinctive from variations of pixel values belonging to the static background [8].

Compared to moving backgrounds occurring in the relatively limited region with a predictable patterns, maintaining the background model with consideration of varying illuminations is a work of the greatest difficulty for outdoor surveillance (see Fig. 1) [9]. More specifically, a sudden illumination change by passing clouds in a short time totally destroys the history of pixel values occluded by casting shadows, which probably leads to the high-level false positive rate as shown in Fig. 1(b). This change also makes the problem of camouflage, that is, moving objects having similar chromatic features to the shadow are hardly detected as foreground. For example, black-suited people are not likely to be detected as foreground in the shadow region even though they move fast in that area. To tackle this problem, intensive researches have been conducted based on various techniques from the field of mathematical engineering as well as computer vision.

This paper is a new attempt to provide a comprehensive review of background subtraction for outdoor surveillance with a special attention on strategies to handle varying illuminations, so-called *illumination-invariant background subtraction* (IIBS). Even though several surveys have been published in literature [10]–[14], to the best of our knowledge, this is the first work to give a in-depth overview focusing on illumination variations for background subtraction. Specifically, this paper explains diverse methodologies devised for IIBS task with a systematic categorization. The performance evaluation using representative methods is detailedly demonstrated based on the benchmark databases and the discussion of the corresponding results is also provided. It is worth noting that this paper can be regarded as a practical guide for beginners or experts by providing qualitative and quantitative results of representative IIBS methods with their processing times. Moreover, since this review efficiently concentrates on a specific and important issue for background subtraction (i.e., varying illuminations) with the corresponding approaches recently proposed, therefore, it is expected to give a great help to develop more advanced algorithms for the high-performance outdoor surveillance system. Finally, this review also provides advantages, weaknesses, and future direction of those background subtraction models for outdoor surveillance. The main contributions of this paper are fourfold:

- In this review, core aspects in illumination-invariant background subtraction systems, e.g., how to formulate the IIBS problem, how to categorize IIBS models, and which factors need to be considered for constructing IIBS systems, are discussed in depth. It is believed that those questions effectively give a general understanding of the IIBS system and inspires newcomers to dramatically contribute to this topic.
- The review presented here pays more attention to turning on the systematic survey based on the vulnerability of background subtraction due to the outdoor dynamics while previous overviews focus on specific methods just in a timeline without any organic connection between them. This is fairly desirable since such problem-driven summarization is helpful to comprehend the key components to build the reliable IIBS system. Moreover, this review appropriately includes recently published works whereas most of methods introduced in previous surveys have become outdated although they are landmark models in this field.
- In contrast to previous comparative studies just listing up individual works, existing methods for IIBS are efficiently grouped according to the proposed taxonomy and various aspects are subsequently analyzed in this paper. The strengths and drawbacks are also discussed in detail.
- Rather than explaining the algorithm of previous approaches without experimental results, the performance comparison is conducted based on benchmark databases, which are collected under varying illuminations, in this review. By comparing the accuracy as well as the processing time, meaningful discoveries have been presented and some insights into the future are shown with constructive discussions.

The rest of this paper is organized as follows. The general procedure of the illumination-invariant background subtraction and a systematic taxonomy are presented in Section II. The comprehensive survey for representative IIBS methods are introduced in Section III. The performance comparison based on benchmark databases is demonstrated and the related discussion is presented in Section IV. Conclusions follow in Section V.

II. ILLUMINATION-INVARIANT BACKGROUND SUBTRACTION: A CONCEPT AND TAXONOMY

A. GENERAL FORMULATION OF IIBS PROBLEM

Many background subtraction models have been proposed to cope with varying illuminations. Although they are designed with different philosophies, most of IIBS methods share a common assumption that the underlying structure of background in a given scene is static (i.e., unchanged) regardless of outdoor dynamics and thus moving objects can be discriminated from background by the simple comparison, which is very similar with the traditional background subtraction [12], [15], formulated as

$$M^k(x, y) = \begin{cases} 1, & \text{if } D(\mathbf{I}^k(x, y), \mathbf{B}_M^k(x, y)) > \tau, \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where $\mathbf{I}^k(x, y)$ denotes the feature vector (e.g., colors, edges, etc.) onto (x, y) pixel position at the k^{th} frame and $\mathbf{B}_M^k(x, y)$ is the background model for the corresponding pixel. $D(\cdot, \cdot)$ and τ denote the distance metric and the thresholding value, respectively. $M^k(x, y)$ thus indicates the motion label (i.e., 1 $\rightarrow$ foreground and 0 $\rightarrow$ background). The overall procedure of IIBS methods is shown in Fig. 2. As can be seen, the background model $\mathbf{B}_M^k(x, y)$ needs to be carefully defined for revealing the underlying structure of background to be illumination-invariant. Notice that the background model $\mathbf{B}$ can be defined in a pixel-level (like this example), block-level, or frame-level. Moreover, it is highly required that the corresponding background model is subsequently maintained with various updating schemes (e.g., online interpolation [16], [17], parametric learning [18], [19], supervised learning [20], [21], etc.) to tolerate the statistical variations by the change of lighting conditions in a short time. In fact, the difference between IIBS methods stems from the way to define and maintain this background model $\mathbf{B}$, and thus our review is conducted by analyzing such strategies of previous approaches. It is thought that this aspect helps readers to understand essential components for constructing robust IIBS frameworks. The detailed categorization of IIBS methods will be demonstrated in the following subsection.

B. A SYSTEMATIC TAXONOMY

In general, it is difficult to reach the universal agreement on the taxonomy of IIBS methods. Nevertheless, IIBS methods can be categorized with different factors for better understanding of a problem. Based on the nature and features employed, IIBS methods can be broadly divided into two main groups: 1) computational and 2) biologically-inspired approaches. Each category can be further classified into various sub-models, which will be introduced in the following Section, with three different attributes, i.e., correlation, scale, and domain. Specifically, the correlation has a connection with whether or not to use the spatial relationship, that is, pixel-based approaches only utilize the current pixel value to define a feature vector while block-based ones combine its neighboring pixels according to the spatial proximity. The scale is related to the metric for the background determination, which can be defined based on the whole image (i.e., global scale) or the sub-image (i.e., local scale). Finally, the domain indicates the characteristics of the feature space (i.e., spatial and frequency space). Figure 3 shows this taxonomy in a hierarchical manner and the corresponding categorization using some representative methods is shown in Table 1, respectively.

First of all, numerous computational approaches have been proposed for IIBS. These approaches aim at suppressing outliers (i.e., background pixels falsely detected as foreground by varying illuminations) based on diverse techniques

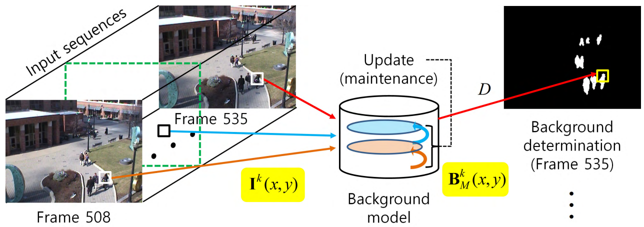


FIGURE 2. Overall procedure of IIBS methods. The background model needs to be consistently updated using previous frames to be illumination-invariant. Notice that the background model can be defined in a pixel-level (this example), block-level, or frame-level.

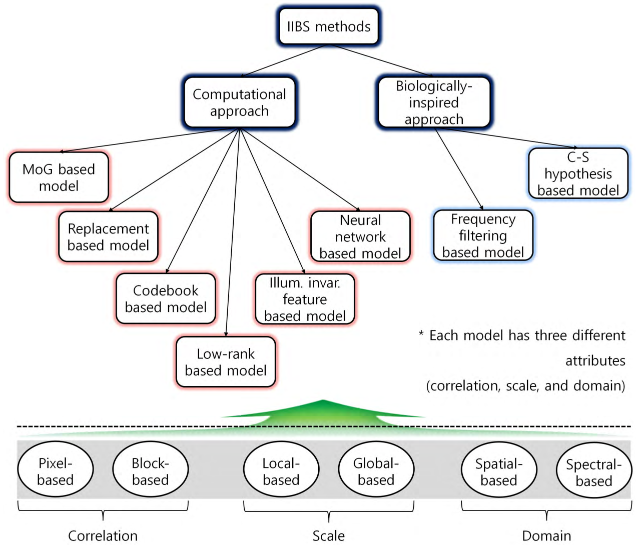


FIGURE 3. A systematic taxonomy of IIBS methods. Notice that two main groups (i.e., computational and biologically-inspired approaches) can be sub-categorized with three different attributes, i.e., correlation, scale, and domain.

TABLE 1. Taxonomy for illumination-invariant background subtraction.

Reference	Category	Correlation	Scale	Domain	Analysis	Key features of methodologies
[22]	CA	Pixel	Local	Spatial	MoG based	MoG models with online updating scheme
[23], [69]	CA	Pixel	Local	Spatial	MoG based	Adaptive setting the number of MoG models with recursive updating scheme
[70]	CA	Block	Local	Spatial	MoG based	PSO-based parameter selection for MoG models
[25], [1], [26]	CA	Pixel	Local	Spatial	Replacement based	Consensus-based pixel replacement for the background maintenance
[27], [28]	CA	Pixel	Local	Spatial	Replacement based	Adaptive determination of the matching threshold and the learning rate for consensus-based background subtraction
[29], [30], [31]	CA	Pixel	Local	Spatial	Codebook based	Color and texture based clustering for background subtraction
[32], [33]	CA	Pixel-Block	Local	Spatial	Codebook based	Codebook generation for the background model based on block color means as well as RGB vector of each pixel
[34]	CA	Pixel	Local	Spatial	Codebook based	Arbitrary-cylindrical color features with random selection of codewords for updating the background model
[38], [40]	CA	Block	Global	Spatial	Low-rank based	Detection of contiguous outliers (i.e., moving objects) in the low-rank representation using RPCA with MRF (c.f. two types of Gaussian kernels for PCA [40])
[39]	CA	Block	Global	Spatial	Low-rank based	Bayesian approach based on the Laplacian distribution for robust matrix factorization (i.e., foreground and background)
[41]	CA	Pixel-Block	Global	Spatial	Low-rank based	Pixel-based motion saliency with the block-based sparse RPCA (two-stage approach)
[45]	CA	Block	Global	Spatial	Low-rank based	Online stochastic framework (i.e., online update) for tensor decomposition
[43], [46]	CA	Block	Global	Spatial	Low-rank based	Efficient schemes for conducting and enhancing RPCA (multilinear PCP [43] and 3D total variation [46])
[47]	CA	Block	Local	Spatial	I-I feature based	Local binary patterns computed from color and intensity spaces as features
[48], [50]	CA	Block	Local	Spatial	I-I feature based	Normalized coefficients computed from the orthogonal transform
[51]	CA	Pixel	Local	Spatial	Neural network based	Reconstruction of the background image based on the RBM learning scheme
[52]	CA	Pixel	Local	Spatial	Neural network based	Subtraction learning based on a single background image and input frames in the CNN framework
[55]	BIA	Block	Local	Spatial	C-S hypothesis based	Dynamic texture model with the KL divergence similarity computation
[56]	BIA	Block	Local	Spatial	C-S hypothesis based	Contrast of ordinal features computed from color and gradient feature spaces (also including the temporal gradient)
[57], [58]	BIA	Block	Local	Spatial	C-S hypothesis based	Contrast of directional coherence in the spatio-temporal domain
[59]	BIA	Block	Local-Global	Spatial	C-S hypothesis based	Superpixel-based spatial and motion contrast with the sparsity of superpixels
[64], [66]	BIA	Block	Local	Spatial	C-S hypothesis based	Self-organizing map based on patterns of incoming pixels
[60], [61]	BIA	Pixel	Global	Spectral	Freq. filtering based	Spectral residual in the frequency domain
[62]	BIA	Pixel	Global	Spectral	Freq. filtering based	Amplitude normalization in the frequency domain with the phase quaternion
[63]	BIA	Pixel	Global	Spectral	Freq. filtering based	Low pass filtering in the log amplitude spectrums

(CA = Computational Approach, BIA = Biologically Inspired Approach, I-I = Illumination-invariant, C-S = Center-surround)

from mathematical engineering and optimization theories. Specifically, most of them attempt to implicitly allow varying illuminations into the background model by exploiting a variety of computer vision techniques, e.g., mixture of Gaussians [22]–[24], pixel replacement [1], [25]–[28], visual codebooks [29]–[34], low-rank representation [35]–[46], illumination-invariant feature representation [47]–[50], neural networks [51], [52], etc. Such dedicated background models efficiently reduce the effect of varying illuminations in background, and it thus leads to the reliable results of background subtraction under outdoor environments. Furthermore, most of them provide a unified framework for handling both varying illuminations and moving backgrounds (e.g., waving trees and rippling water), which is useful for a wide range of outdoor surveillance scenarios.

As alternatives to computational approaches, biologically-inspired solutions have been studied. These approaches attempt to mimic the human visual system (HVS), which has an ability to quickly grasp the relevant regions even in a complex scene without any training. In this category, they formulate the problem of IIBS as detecting salient regions under varying illuminations. Since the contrast is generally thought as the most important factor to contribute such visual saliency based on the neurobiological experiments [53], [54], most of them have been devised by exploiting the center-surround hypothesis, i.e., the relevant context in a given scene is not determined by the absolute value of the visual information, but rather the difference between a current point and its surroundings (i.e., contrast) [56]. In this point of view, moving objects showing distinctive motions can be regarded as most salient regions in a given scene and thus expected to be successfully detected regardless of dynamic changes of background (e.g., varying illuminations and moving backgrounds). That is, any additional task for considering various factors to maintain the background model under outdoor environments is not required in this framework. This saliency-based schemes have been successfully applied to IIBS methods with the contrast of spatial features [55]–[59] and filtering of the frequency coefficients [60]–[63]. On the other hand, some methods have adopted the topological relationship-based visual attention mechanism, which preserves the distribution of the background model using neighbor connections on the grid network, for maintaining the background model [64]–[66]. It is noteworthy that most of saliency-based background subtraction approaches (especially using the spectral coefficients) perform very fast due to their nature taking HVS off and can be directly embedded in the camera.

In the following Section, we provide methodologies of diverse approaches for illumination-invariant background subtraction in detail, based on the taxonomy presented above.

III. ILLUMINATION-INVARIANT BACKGROUND SUBTRACTION: METHODOLOGIES

A. COMPUTATIONAL APPROACHES FOR IIBS

1) MIXTURE OF GAUSSIANS BASED MODELS

Over the last years, pixel-based methods have been increasingly proposed due to the ability for accurately preserving the shape of moving objects. Among them, the mixture of Gaussians (MoG) has been widely employed up to now. Most notably, Stauffer and Grimson [22] proposed to formulate the distribution of pixel values over the time as MoG. In addition, they attempted to update MoG models defined at each pixel position based on a simple clustering technique with the online interpolation. The multimodal nature of MoG-based approaches leads to the reliable results in many practical situations [67], [68] and still has gained a lot of interests enough to encourage researchers for continuing to enhance the MoG-based schemes. Specifically, Zivkovic [23] proposed to adaptively determine the number of Gaussians across frames for alleviating variations by globally casting shadows and further utilized the recursive updating scheme to allow illumination changes into the Gaussian model efficiently [69]. Similarly, White and Shah [70] proposed to exploit the particle swarm optimization technique to find the parameters of Gaussian models appropriately, which is effective to cope with varying illuminations. The downside of MoG-based approaches lies in the variance of Gaussians, that is, variations of pixel values due to illumination changes are hardly estimated in advance and thus cannot be set as the optimal value for all the outdoor environments.

2) PIXEL REPLACEMENT BASED MODELS

In contrast to delicately design the probability density function for the background model (e.g., MoG-based approaches), the most confident value among last observed pixel values can be selected as the background model at each pixel position. This nonparametric scheme may avoid problems driven by suboptimally assumed density models. First of all, Wang *et al.* [25] applied the sample consensus (SACON) method to background subtraction, which simply counts the number of times previous samples agreeing (i.e., small distance) with the current sample. Notice that if the counting number is larger than a pre-defined threshold, the current pixel is determined as background. Barnich and Droogenbroeck [1] refined this simple scheme by randomly selecting which values to substitute from the background model. Their visual background extractor (so-called ViBe) is, thus, able to be adapted to varying illuminations within a relatively short period of time. Moreover, they propagated the value of the background model into its surroundings especially for preserving background regions temporally hidden by foreground. Once the color feature of the current pixel leads to update the set of samples for maintaining and representing the background model, this pixel is also employed to update its neighbor regions. Since such consensus based approaches are conceptually simple and easy to implement, many variants have been proposed in literature. Droogenbroeck and Paquot [26] proposed some post-processing techniques to polish the result of ViBe, for example, inhibition of propagation around internal boundaries to prevent the split effect in a single object. Hofmann *et al.* [27] proposed to adaptively set the decision threshold and learning parameters across frames by estimating background dynamics (i.e., analyzing pixel-level motion patterns of background), which are the core of background determination and update while keeping the algorithm structure of ViBe. With simple modifications, they efficiently improve the performance of ViBe. St-Charles *et al.* [28] tried to incorporate the local binary similarity patterns (LBSP) [71], which can be regarded as a counterpart to local binary patterns (LBP) [72], into the sample consensus-based updating framework for IIBS. Furthermore, they modified the feedback scheme [27] to adaptively determine both the distance threshold and the learning rate by measuring the motion entropy at each pixel position. Even though methods mentioned above perform very fast and efficiently remove dynamic backgrounds, most of them require the additional cache for storing "recently observed frames", which is hardly allowable in the embedded surveillance system.

3) CODEBOOK BASED MODELS

The concept of the visual codebook has been widely employed in computer vision fields, for example, scene recognition [73], [74], object detection [75], [76], action recognition [77], [78], etc. Since the codebook is capable of encoding the diversity of a given class (i.e., background) without any supervisory procedure (i.e., unsupervised clustering), it can be effectively applied to achieve IIBS. Specifically, Kim *et al.* [29] proposed to generate a codebook using sample background RGB values with 6-tuple features including minimum and maximum color values, statistical information of each codeword, etc., for a given video sequence. For each pixel in the current frame, the distance from cluster means (i.e., codewords) is computed to determine whether the feature vector defined for the corresponding pixel is successfully matched with the codebooks or not (if matched, then that pixel belongs to background). Other codebook-based approaches basically follow this scheme since it allows us to successfully capture the structural background variations, however, strategies to build up the codebook features used are different. Doshi and Trivedi [30] proposed to combine the color codebook [29] with the shadow suppression model, so-called the hybrid cone-cylinder model, for reducing the effect of illumination changes. Pal *et al.* [31] proposed a technique of the codeword spreading across layer boundaries to handle background variations. To do this, codewords need to be distributed along the boundary of two neighbor regions and the codebook is made up with codewords from the other side of the boundary. Guo *et al.* [32] proposed a hierarchical scheme with the combination of block-based and pixel-based codebooks. They utilized 12 color means computed from each block to conduct codebook matching in a similar way of [29], which roughly enables to extract foreground regions with a simple distance metric, and subsequently applied the pixel-based codebook to refine the result of background subtraction. Zeng *et al.* [34] exploited a modified arbitrary-cylindrical color model, which is controlled by the center, direction, and two bounds, to construct a more reliable codebook. They further developed a equal-qualification updating scheme based on the random selection of codewords to be robust to varying illuminations. Even though codebook-based approaches have been popularly employed for IIBS, it is still difficult to determine the optimal size of the codebook.

4) LOW-RANK BASED MODELS

In a given video frames, background can be formulated as a low-rank matrix while moving objects are detected as outliers. The robust solution of this representation is driven by Torre and Black [79] with a pioneer work called robust principal component analysis (RPCA), and applied for background subtraction by Candes *et al.* [37]. Specifically, they utilized principal component pursuit (PCP) as a solver for RPCA [37]. In this framework, gradual changes (e.g., varying illuminations) over time can be successfully approximated by the corresponding low-rank subspace while moving objects consists of correlated sparse components. Due to its promising results regarding to background dynamics, various mathematical techniques to solve the factorization problem have been actively studied. In particular, to improve the performance of background subtraction, the spatially-contiguous property of pixels has been recently integrated into RPCA based on Markov random fields (MRFs) and other smoothing schemes. Zhou *et al.* [38] adopted the convex optimization technique for approximation of the low-rank matrix while applying the combinatorial optimization to estimate the outlier support. Compared to other formulations of RPCA, they estimate the outlier support (i.e., foreground) explicitly and thus the contiguous prior between pixels can be naturally incorporated into RPCA using MRF in this method (named as DECOLOR), which significantly improves the performance of background subtraction under varying illuminations. Wang and Yeung [39] proposed a full Bayesian approach for the robust matrix factorization. To this end, they adopted a Laplacian mixture model as the likelihood and designed an efficient sampling scheme by using the hierarchical Laplacian distribution. Javed *et al.* [40] tried to improve the performance of IIBS with the spatially continuous constraint in a similar way of DECOLOR [38]. Differently from previous approaches, they decompose a given frame by using two different Gaussian kernels, which yield Gaussian and Laplacian images, and subsequently apply the online PCA for background modeling. Yao *et al.* [41] proposed a two-stage framework for background subtraction. More concretely, they first conduct the low rank and structured sparsity decomposition to get candidates of foreground regions, which are subsequently filtered out by the motion saliency map. The block-based sparse RPCA is finally employed to decide the size and location of moving objects.

Since video surveillance is generally conducted over multiple frames (i.e., image volume), some researchers have exploited IIBS methods based on the analysis of the tensor space to improve the accuracy of background subtraction even requiring more complex computational procedures. Similarly, most of tensor-based methods have focused on how to generalize a given volume by utilizing the RPCA optimization scheme. Li *et al.* [43] extended the PCP algorithm in a multilinear way for the optimal tensor decomposition, which is formulated as a convex optimization with relaxation techniques efficiently improving the performance of background subtraction. Recently, Sobral *et al.* [45] have proposed an online stochastic framework for tensor decomposition of a given video sequences, which efficiently reduces the computational burden for matrix factorization. Basically, they also include the background dynamics into the low-rank component by the stochastic optimization applying on each mode of the tensor. Cao *et al.* [46] utilized the 3D total variation to enhance the spatio-temporal continuity of foreground regions in a tensor framework, which efficiently suppresses the effect of casting shadows. They decompose video frames into backgrounds and moving objects with such tensor-based RPCA scheme.

In contrast to other previous models, the low-rank representation has the remarkable improvement for IIBS by implicitly integrating the effect of illumination changes into the low-rank space, however, many parts of low-rank based approaches are time-consuming and require the additional memory spaces for the batch-based processing. Notice that other optimization techniques also can be applied to IIBS for improving the processing speed [80].

5) ILLUMINATION-INVARIANT FEATURE BASED MODELS

Methods in this category aim at designing a specific descriptor to represent the underlying structure of background regardless of outdoor dynamics. Rather than implicitly tolerating illumination changes into the background model, these approaches explicitly make a feature to be illumination-invariant, which is adopted to define the background model. It is worth noting that such a good feature to reliably describe the background statistics even under complex real-world situations makes the procedure of the background update simple. To this end, the local binary pattern (LBP), which is one of the most representative descriptors, has been widely used [72]. Since this LBP descriptor checks the relative difference between spatially neighboring pixels not the absolute values of each pixel, illumination changes can be efficiently overcome. Thanks to its simple computation, the LBP descriptor can be easily extended to the spatio-temporal domain for IIBS. For example, Chua *et al.* [47] combined local color patterns with texture ones (i.e., original LBP) to enhance the performance of IIBS in the homogeneous background. In the following, the background model is simply updated with these two histogram features using the online interpolation. Zhang *et al.* [48] proposed the ratio edge as a feature to be employed for detecting moving cast shadows, which is defined based on the relationship of coefficients obtained from orthogonal transforms, e.g., discrete cosine transform (DCT), discrete Fourier transform (DFT), Haar transform, etc. Kim and Kim [50] analyzed the effect of varying illuminations using the statistical information of the intensity lattice, that is, the variation of pixel values on the small local region mainly occurs by illumination changes and thus can be represented by the dominant energy in the corresponding region. Based on this observation, they defined the unit brightness level by simply normalizing the coefficients of the singular value decomposition (SVD) to the largest one, performed at each pixel position. Notice that the processing speed of those methods strongly depends on the complexity of feature descriptors (unfortunately, most of high-performance models works slowly).

6) NEURAL NETWORK BASED MODELS

Finally, some researchers have attempted to apply the learning method for the maintenance of the background model by using the deep neural network (DNN). Compared to other approaches explained above, the deep learning-based background modeling has an ability to contain a wide range of variations in its layer structure and thus greatly estimates background regions even under various dynamics of outdoor environments. Since DNN-based IIBS is in its early stage, corresponding studies have not progressed actively. Rafique *et al.* [51] attempted to reconstruct the underlying structure of background using the restricted Boltzmann machines (RBM). They put the background samples into the receptive field of the RBM framework and reconstructed the background region of incoming frames by using the generative path of RBM. Braham and Droogenbroeck [52] proposed to adopt the convolutional neural network (CNN), which has shown surprising results in a wide range of computer vision fields (e.g., image recognition, object detection, etc.), for background subtraction. In this scheme, authors used a single background image, which does not include any moving object, and a scene-specific training samples to learn how to subtract the background region from the input frame even captured under varying illuminations. While neural network-based approaches do not require the complex modeling procedure for background subtraction, the number of training samples is not enough to reliably train CNN since those are obtained online from a given input sequences during a short time. Nevertheless, due to the power of the deep learning techniques, many researchers now start to give considerable attentions on DNN-based IIBS [81].

B. BIOLOGICALLY INSPIRED APPROACHES FOR IIBS

1) CENTER-SURROUND HYPOTHESIS BASED MODELS

As explained in the previous Section, the contrast is the most influential component to detect salient motions in a given scene. Methods in this categories assume that such salient motions are generated by moving objects while other dynamics are dropped into non-salient regions (i.e., background). In this point of view, Mahadevan and Vasconcelos [55] employed the dynamic texture model as their features and computed the discriminant center-surround contrast based on the KL divergence. Kim *et al.* [56] combined the spatial ordinal features and the temporal gradients (i.e., motion features computed from frame differencing) into the center-surround framework. Notice that the temporal coherence, which is defined as the energy concentrated on the dominant direction of the gradient distribution obtained from the difference image, also brings the performance improvement of detecting salient motions [57]. They further extended the concept of the coherence to the spatio-temporal domain using the structure tensor, which shows the competitive results compared to traditional methods for background subtraction [58]. Liu *et al.* [59] proposed an adaptive fusion method to integrate spatial and temporal saliency into a single saliency map. Specifically, they first computed the superpixel-based global contrast and the spatial sparsity of superpixels. The motion distinctiveness of superpixels is subsequently combined with the spatial saliency to derive the spatio-temporal saliency map. Since this method is based on superpixels, the shape of moving objects is well preserved even under complex background. On the other hand, Maddalena and Petrosino [64] proposed a novel approach automatically adapting to illumination changes in a self-organizing manner without any prior knowledge. This method (called SOBS) allows to preserve the topological relationship of input patterns and thus works robust to variations of background including gradual illumination changes, moving backgrounds, and camouflage. Patterns of incoming pixels are mapped to the neuron whose set of weights is most analogous to the corresponding pattern and weight vectors belonging to the neighbor region are updated. It often has been slightly modified for other applications, for example, pan-tilt-zoom cameras [66]. The motivation of those models is quite convincing for background subtraction, however, most of them still suffer from casting shadows making the large contrast between neighbor regions.

2) FREQUENCY FILTERING BASED MODELS

As a pioneer work, Hou and Zhang [60] firstly considered the frequency domain to find salient regions. Based on extensive experiments, it is observed that the spectral residual tightly corresponds to the visual saliency. Inspired by this simple and surprising result, Cui *et al.* [61] extended the spectral residual to the temporal domain for background subtraction. They decomposed the video volume into horizontal-temporal and vertical-temporal planes, and applied the spectral residual to each slice. Based on the majority vote, background regions are clearly extracted from outdoor scenes. Guo and Zhang [62] introduced the phase quaternion Fourier transform (PQFT) to consider the motion information as well as color features, and applied it to the outdoor monitoring. It is noteworthy that such filtering techniques in the frequency domain enables background subtraction to work extremely fast ($<$ 1ms), which is desirable for embedded surveillance systems. Recently, Li *et al.* [63] have shown that the low pass filtering in the log amplitude spectrum is actually relevant to the salient regions, rather than the spectral residual. By building the spectrum scale space with such filtering results, arbitrary types of salient regions can be greatly detected. Notice that the motion information can be easily incorporated into the hypercomplex Fourier transform scheme [63] for IIBS. Even though those methods perform very fast due to the power of FFT operations, the boundary of moving objects tend to be blurred by filtering and resizing operations, which may make the performance drop in further applications, e.g., tracking and recognition.

IV. EXPERIMENTAL RESULTS

In this Section, we have demonstrated the performance of background subtraction via various algorithms introduced in the previous Section based on two benchmark databases particularly constructed for varying illuminations, which are PETS2001 [82] and OTCBVS [83]. These two databases are publicly available and have been widely used for the performance evaluation of background subtraction in outdoor environments. Specifically, the PETS2001 database (category: data3) is relatively captured at a distance and contains the global illumination changes. On the other hand, the OTCBVS database includes casting shadows frequently occurring in a short time, which yields the structural change of local regions. Some samples from two databases are shown in Fig. 1.

A. QUALITATIVE EVALUATION

For the performance evaluation of IIBS, we employ total 12 methods among diverse models introduced in previous Section, which are A-MOG [23], ViBe [1], SuBSENSE [28], BGFG [29], DECOLOR [38], OSTD [45], SILK [80], TC-LBP [47], IISC [50], SC-SOBS [65], PQFT [62], and DTSD [56]. Notice that at least one method for each sub-category is selected except the neural network based approach. Some results of background subtraction using PETS2001 and OTCBVS databases are shown in Fig. 4 and 5, respectively.

First of all, SuBSENSE, DECOLOR, SILK, and IISC approaches show quite a reliable results of background subtraction under globally varying lighting conditions compared to other methods as shown in Fig. 4(e), (g), (i), and (k). In contrast, some methods, e.g., A-MOG and BGFG, yield high-level false positives when illuminations start to change, for example, from cloudy to sunny (see the second row and the third row in Fig. 4(a)). On the other hand, biologically-inspired approaches attempt to detect other visually attractive regions, e.g., parked cars, which often belong to the static background. They also tend to more emphasize the smaller object even with the multiple moving ones. In the OTCBVS database, clouds pass fast in a short time while yielding structured shadows, which leads to the significant variations from the initial background status. In particular, unwanted edges by the casting shadow makes the problem more intractable as shown in Fig. 5(a). Under such challenging conditions, most of background subtraction methods frequently fail to detect moving objects accurately. Specifically, it is easy to see that low-rank based approaches are vulnerable to the abrupt change of lighting conditions (see Fig. 5(g), (h), and (i)) whereas they successfully tolerate global changes of illuminations as shown in Fig. 4. MoG-based and codebook-based methods also suffer from casting shadows locally occurring in a short time. They yield uneven results of background subtraction according to illumination variations in local regions. Even though the LBP-based model works successfully in global variations of lighting conditions, it fails to construct the robust background model under the abrupt change of illuminations. In contrast, pixel replacement based approaches (e.g., ViBe and SuBSENSE) show consistent and reliable results throughout the entire sequence as shown in Fig. 5(d) and (e). It should be emphasized that the IISC method, which is based on illumination-invariant features, provides robust results of background subtraction regardless of globally or locally varying illuminations (see Fig. 4(k) and 5(k)).

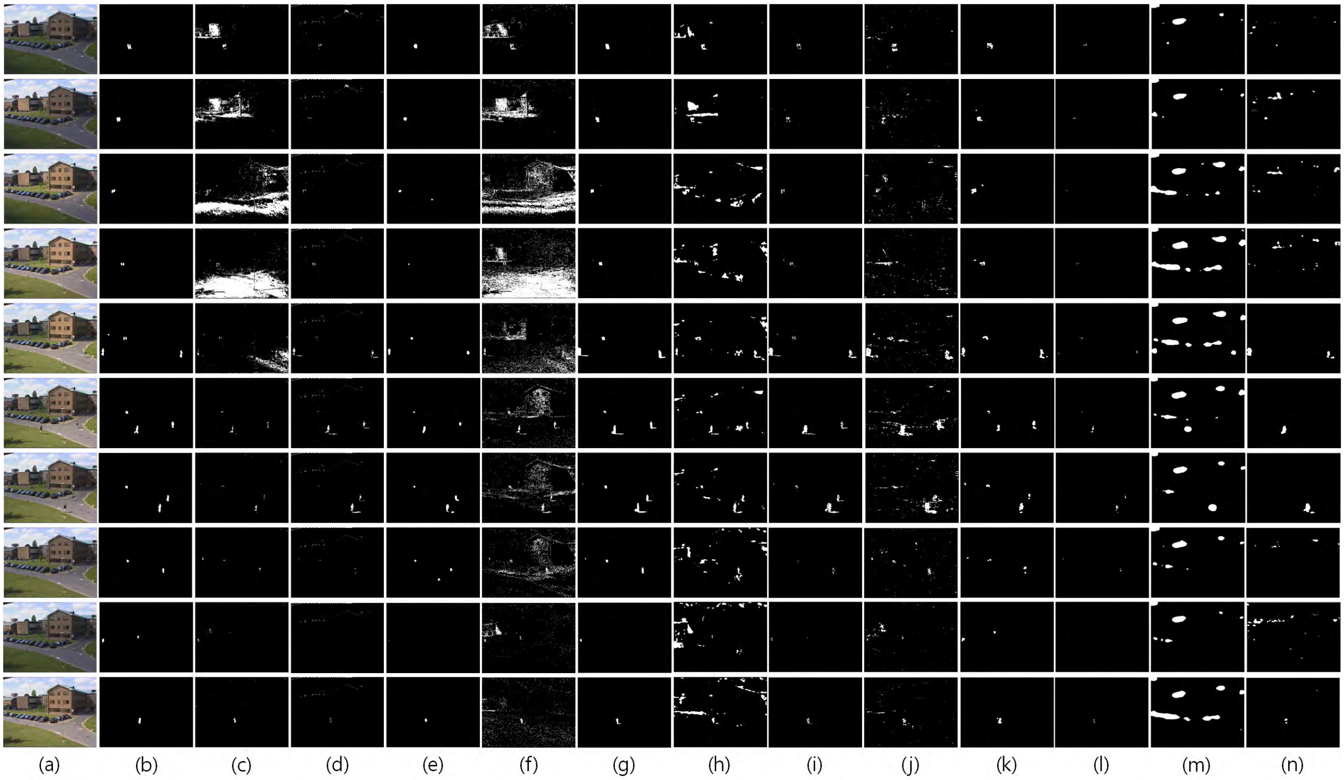


FIGURE 4. Results of background subtraction in the PETS2001 database. (a) Input frames, (b) Ground truth, (c) A-MOG [23], (d) ViBe [1], (e) SuBSENSE [28], (f) BGFG [29], (g) DECOLOR [38], (h) OSTD [45], (i) SILK [80], (j) TC-LBP [47], (k) IISC [50], (l) SC-SOBS [65], (m) PQFT [62], and (n) DTSD [56]. Notice that illuminations globally change across the whole range of a given scene as shown in (a).

TABLE 2. Performance comparison by F-scores in PETS2001 and OTCBVS databases.

PETS2001	A-MOG	ViBe	SuBSENSE	BGFG	DECOLOR	OSTD	SILK	TC-LBP	IISC	SC-SOBS	PQFT	DTSD
Recall	0.430	0.234	0.515	0.602	0.772	0.808	0.526	0.607	0.671	0.129	0.308	0.372
Precision	0.019	0.069	0.816	0.011	0.635	0.053	0.590	0.129	0.459	0.933	0.015	0.063
F-score	0.036	0.107	0.632	0.022	0.697	0.099	0.556	0.212	0.546	0.226	0.029	0.108
OTCBVS	A-MOG	ViBe	SuBSENSE	BGFG	DECOLOR	OSTD	SILK	TC-LBP	IISC	SC-SOBS	PQFT	DTSD
Recall	0.671	0.620	0.898	0.784	0.984	0.869	0.688	0.652	0.890	0.727	0.463	0.356
Precision	0.308	0.625	0.494	0.109	0.142	0.153	0.171	0.134	0.418	0.389	0.180	0.438
F-score	0.423	0.623	0.638	0.193	0.248	0.259	0.274	0.222	0.569	0.507	0.259	0.386
Avg. F-score	0.230	0.365	0.635	0.108	0.473	0.179	0.415	0.217	0.558	0.367	0.144	0.247

B. QUANTITATIVE EVALUATION

In this subsection, we conduct the performance comparison quantitatively based on the manually-labeled ground truth in both databases (see Fig. 4(b) and 5(b)). To do this, we randomly select total 100 frames from each video, which are spaced throughout the entire sequences. For the performance evaluation, the F-score is computed since it efficiently considers the balance between the hit rate of ground truth (i.e., recall) and the false positives (i.e., precision) as follows [84]:

$$F = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision}. \tag{2}$$

Notice that the F-score is suitable for measuring the capability to suppress falsely detected pixels as background while preserving the shape of moving objects, and thus employed in most of previous studies [14], [50].

The results of the performance comparison are shown in Table 2, Fig. 6, and 7, respectively. Based on F-scores in Table 2, it is thought that SuBSENSE, DECOLOR, and IISC methods reliably work under varying illuminations compared to other approaches. Those have an ability to preserve the shape of moving objects while successfully suppressing the effect of casting shadows. It is noteworthy that relative low F-scores even with high recall rates are often shown by the large amount of falsely detected pixels as background in other models, e.g., BGFG, OSTD and TC-LBP. The abrupt change of the structural information, e.g., occlusion by casting shadows, is hardly decomposed into the low-rank space, which leads to the notable drop of precision as shown in Table 2, even though they are robust to the global change of lighting conditions occurring in the PETS2001 database. Biologically-inspired approaches often fail to uniformly

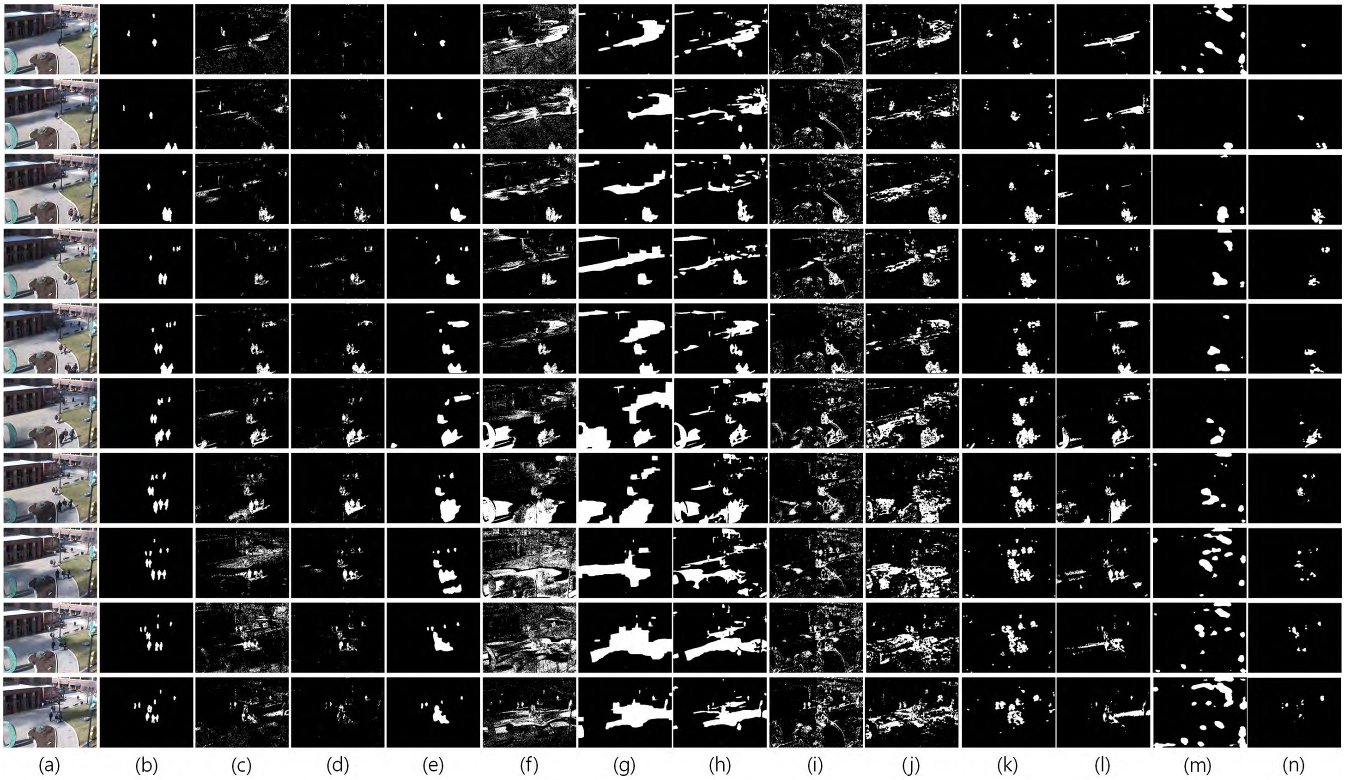


FIGURE 5. Results of background subtraction in the OTCBVS database. (a) Input frames, (b) Ground truth, (c) A-MOG [23], (d) ViBe [1], (e) SuBSENSE [28], (f) BGFG [29], (g) DECOLOR [38], (h) OSTD [45], (i) SILK [80], (j) TC-LBP [47], (k) IISC [50], (l) SC-SOBS [65], (m) PQFT [62], and (n) DTSD [56]. Notice that this database contains the abrupt change of lighting conditions by the casting shadows, which fast pass through buildings.

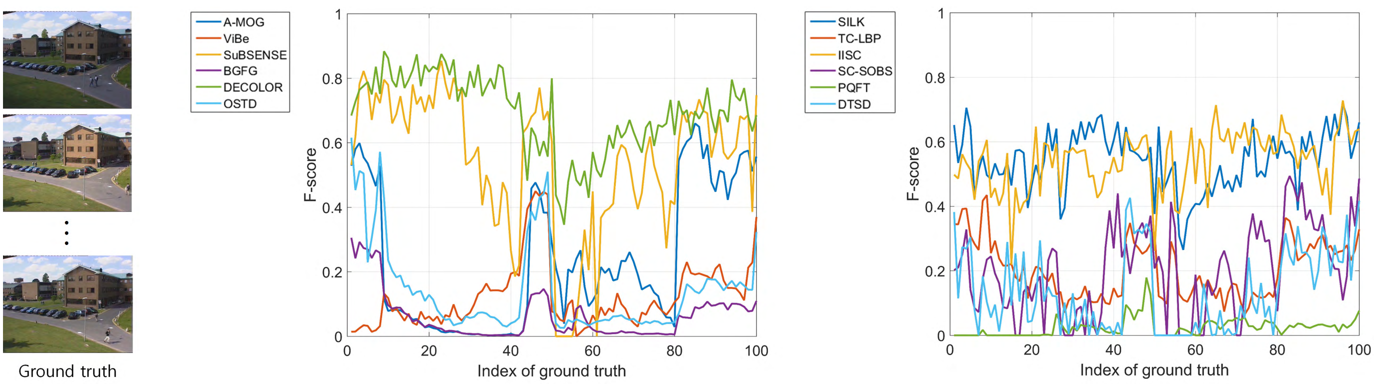


FIGURE 6. Performance variation in ground truth frames of the PETS2001 database. Notice that most of approaches show the performance drop between 40th and 60th samples in the ground truth, which contain varying illuminations changing like this: sunny→cloudy→sunny in a short time.

capture multiple moving objects as mentioned above. Notice that, unlike PQFT and DTSD, SC-SOBS method is quite a robust to varying illuminations owing to the usage of the spatially-connected topological relationship. On the other hand, the codebook based method is vulnerable to the big difference of color values since the codebook is generally defined by using several frames in the beginning part of a video sequence and slowly updated. Figures 6 and 7 show variations of F-scores according to each frame of ground truth in both databases. For the reliable outdoor surveillance, it is desirable for the algorithm of background subtraction to maintain the consistent performance regardless of outdoor dynamics, however, most of methods suffer from the unstable results strongly dependent on lighting conditions at that moment as shown in Fig. 6 and 7. Notice that the IIBS performance varies dramatically in the PETS2001 database due to illumination changes occurring in the overall region (i.e., large amount of pixels are changed simultaneously) compared to that of the OTCBVS database. According to the average of F-scores, we demonstrate more examples of

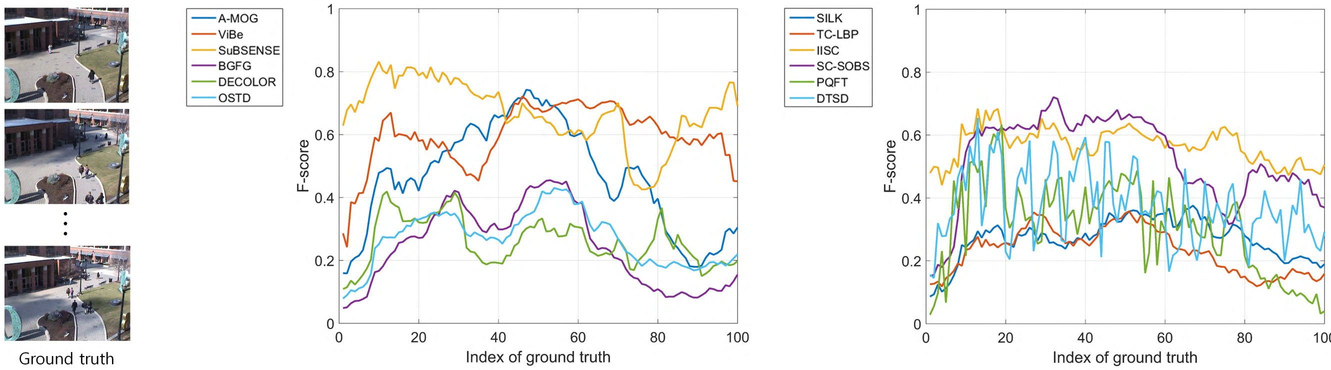


FIGURE 7. Performance variation in ground truth frames of the OTCBVS database. Notice that most of approaches show the performance drop between 60th and 80th samples in the ground truth, which contain abrupt changes by passing clouds leading to the structural difference in background.

TABLE 3. Performance comparison by the processing speed.

	A-MOG	ViBe	SuBSENSE	BGFG	DECOLOR	OSTD	SILK	TC-LBP	IISC	SC-SOBS	PQFT	DTSD
Speed (fps)	76.92	66.67	10.87	142.86	0.29	2.27	2.67	22.22	6.25	55.19	109.89	23.26
Implement	C++	C++	C++	C++	Matlab	Matlab	Matlab	C++	C++	C++	Matlab	C++

background subtraction by SuBSENSE and IISC approaches as shown in Fig. 8 for better understanding with regard to the IIBS performance.

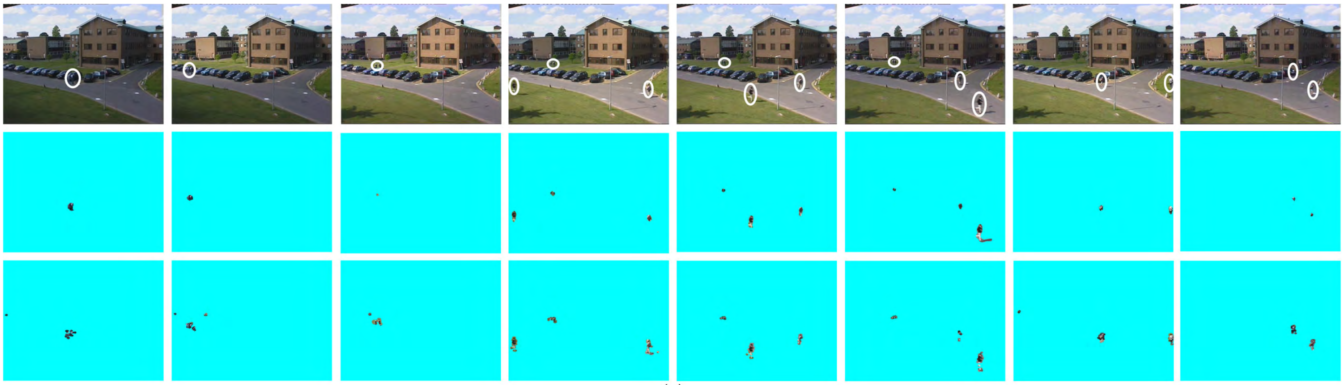
C. ANALYSIS OF THE PROCESSING SPEED

For the performance comparison, all the methods are evaluated on the PC with Intel i7 3.4GHz and 16GB RAM. The framework for our experiments is constructed by utilizing both the Visual Studio 2015 C++ and Matlab. Notice that source codes available in the public domain are employed without any significant change for our experiments. The comparison of the processing speed is shown in Table 3. As can be seen, the processing speed of low-rank based approaches is relatively slow since most of them employ the batch processing with the complex mathematical optimization while pixel-based and biologically-inspired approaches work very fast. For the success of outdoor surveillance, IIBS algorithms need to work fast enough to be embedded into mobile devices. Therefore, the real-time processing while keeping the performance of IIBS is still an open problem in the field of computer vision.

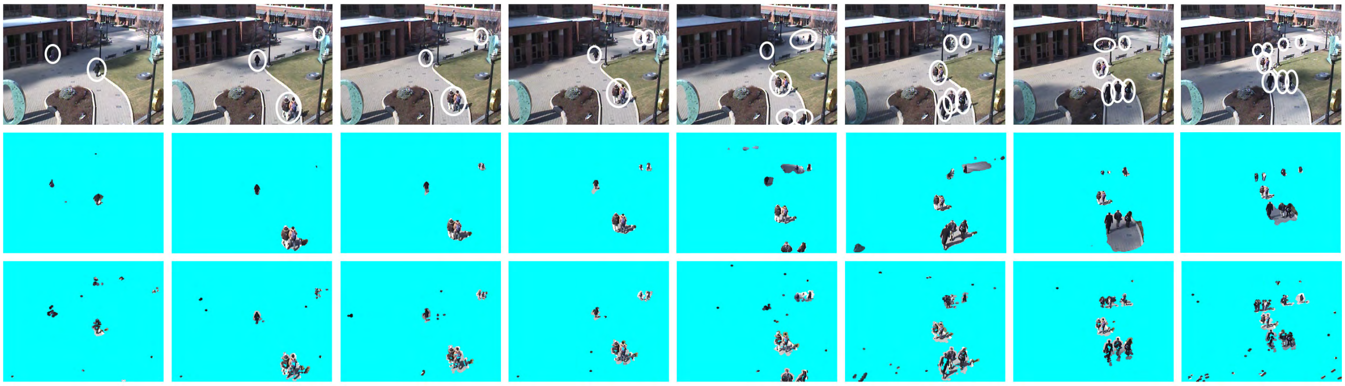
D. SUMMARY AND DISCUSSION

Beyond indoor monitoring under constrained environments, intelligent surveillance systems now start to be deployed into outdoor environments. It is highly desirable to efficiently control the wide range of public spaces and promptly protect people from terrorism. This also enables the crowd analysis and motion forecasting more robustly. However, outdoor dynamics, e.g., varying illuminations and moving backgrounds, make it difficult to apply such useful applications to a wider fields. In particular, varying illuminations makes background subtraction more intractable since it often changes the structural information of a given scene in a short time without providing any prior knowledge. A great amount of researches has been carried out to resolve this problem for outdoor surveillance and some of plentiful approaches are introduced in this paper. Based on various experimental results shown in previous subsections, pros and cons of each category can be summarized as follows:

- **MoG based models** : In general, MoG based methods extract moving objects from background with the high recall rate, however, they are vulnerable to the short-term change by varying illuminations, which yields quite a large amount of falsely detected pixels. This is because the learning rate is not adaptive to the change speed of lighting conditions and thus Gaussian models cannot be reliably updated to tolerate such abrupt changes. Nevertheless, these models still have been popularly employed for outdoor surveillance due to the fast processing speed. Therefore, the usability of MoG based methods can be improved with various supporting strategies, e.g., adaptively updating the background model.
- **Pixel replacement based models** : Since these methods conduct sampling based on a set of pixels stored during a specific time period (e.g., 20 frames) and update the background model with the most similar pixel by the replacement, they are quite a robust to the abrupt change of lighting conditions and also perform fast. Furthermore, algorithms for shadow detection can be straightforwardly applied to this replacement schemes, which efficiently improves the performance of IIBS. Even though these models work robustly under varying illuminations, they require additional memory spaces to store the history of pixel values, which places a heavy burden on the embedded surveillance system.
- **Codebook based models** : Since the codebook, which represents the background model as several codewords, is generally defined at the beginning part of a given video with initial frames (e.g., 100 frames), it hardly



(a)



(b)

FIGURE 8. More examples of background subtraction with SuBSENSE [28] (middle) and IISC [50] (bottom) approaches. (a) Results on the PETS2001 database. (b) Results on the OTCBVS database. The white ellipses in the original frame (top) indicate moving object. Notice that these methods provide reliable results of background subtraction in most parts of given sequences.

considers the large difference of color values at each pixel position by the casting shadows, which occurs in arbitrary parts of the video sequence. It is worth noting that the philosophy of codebook based models is similar with that of MoG based approaches in terms of using the distance metric based on pre-defined color clusters. Therefore, the codebook can be updated like Gaussian models, however, it still does not respond well to the rapid change of color values due to similar reasons of MoG based approaches. One important advantage of codebook based methods is that the spatial relationship between neighbor pixels can be efficiently encoded into the codebook. To improve the performance of codebook based IIBS methods, strategies for associating such local information with the codebook and constructing the codebook with a capability to cover a wide range of variations need to be actively explored.

- **Low-rank based models** : Most of methods in this category attempt to decompose a given frame into the low-rank background model and sparse outliers (i.e., moving objects). To do this, robust PCA-based schemes have been most widely employed with various optimization techniques. Since these methods implicitly outdoor dynamics (moving clutters as well as varying illuminations) from the static background by matrix factorization techniques, the accuracy of IIBS is acceptable in most scenarios. Therefore, many algorithms for solving the decomposition problem efficiently have been exploited [85]–[88]. However, most of them require a batch processing with complex operations for the large-sized matrix, which is time-consuming (see Table 3). In order to apply low-rank based approaches to the limited resource environments, more advanced engineering skills, e.g., online decomposition and one-shot decomposition, need to be developed.
- **Illumination-invariant feature based models** : Since most of researchers focus on designing the feature descriptor to efficiently reveal the underlying structure of background rather than directly using pixel colors, they have a loss in the processing speed while showing the better IIBS performance compared to other pixel-based fast algorithms. Such features can be easily combined with any type of background updating schemes without severe modification, which enables this feature based algorithm to be adaptively applied according to the resource of the target system. Moreover, other techniques in the field of computer vision, for example, color consistency [89], face normalization [90], etc., can

be efficiently adopted to devise the feature descriptor, which are very helpful for improving the performance of IIBS. However, as mentioned before, the processing time of feature based models is proportional to the complexity of feature computations, and thus they need to be carefully designed for generality.

- **Neural network based models** : With rapid development of deep learning techniques, some of methods as explained in Section III start to employ various structures of the neural network for background subtraction. In particular, the underlying structure of background can be trained with a few number of nodes in the hidden layer. In addition, the subtraction procedure itself is formulated as training of a given network with the convolutional operator. Thanks to the abstraction ability, such neural network based approaches begin to show their greatness in suppressing the effect of outdoor dynamics although only few number of models have been introduced in literature. The performance of IIBS by neural network based approaches is comparable to other high-performance methods, however, these models require a large number of training samples to stably train the network, which is carried out offline. Therefore, it is hard to update the background model (i.e., weights of the neural network) by using current frames of a given video in an online manner. To bridge this gap, incremental learning structures for IIBS applications need to be considered in depth.
- **Center-surround hypothesis based models** : As compared to previous models mentioned above, these method focus on the contrast defined by computing the difference of features between center and its surrounding regions to discriminate moving objects from background. Due to the pooling procedure (e.g., average, maximum, etc.), small variations occurring by illumination changes can be efficiently suppressed. In this category, the additional scheme for updating the background model is not required. However, uniformly highlighting all the moving objects cannot be easily achieved due to the contrast normalization over the whole region of a given frame (i.e., small contrast frequently fades out). Moreover, the processing speed is highly dependent on the complexity of features used for computing the contrast. Even though DTSD method works fast as shown in Table 3, some other ones based on the subspace analysis operate slowly, e.g., 1.72 fps [57] and 15 fps [58], which is insufficient to meet the real-time processing criteria in the embedded surveillance system.
- **Frequency filtering based models** : These models are conceptually simple and easy to implement. The processing speed is also very fast due to the global filtering based on FFT (fast Fourier transform) in the spectral domain. Even though they show the good performance to reduce the effect of outdoor dynamics, the small amount of motions is highly likely to be filtering out since the corresponding regions do not make a notable temporal difference from background, which leads to the failure of capturing the small size of moving objects as shown in Fig. 4(m). Moreover, since such filtering process for amplitude coefficients in the spectral domain often yields the edge detection-like results [63], the boundary regions generated by the casting shadow in background are hardly suppressed. To get the most out of fast speed while supplementing weak points, it is helpful to employ the multiscale framework for extracting all the moving objects accurately from background.

In addition to the fundamental analysis mentioned above, practical studies for IIBS in outdoor environments should not be forgotten. Hardware devices for intelligent surveillance systems rapidly progress and various techniques in computer vision continue to merge with such devices. It is noteworthy that some of advances can provide a key solution to develop the efficient and robust IIBS methods. For example, matrix-based complex computations are not suitable for the embedded surveillance system due to the limited computing resources just a few years ago. However, nowadays, various embedded boards, which are capable of providing the computing performance similar to the high-end PC, have been widely employed at a very low price. The community for IIBS also need to engage in developing new methods particularly regarding the neural network based approaches. As shown in a wide range of image (or video) based applications, it is expected that deep learning techniques have become a great help in resolving problems for IIBS. Some of researchers already start to employ the deep learning scheme for IIBS, however, seed components for background subtraction, e.g., online-updatable structures, still are not considered. In summary, even though a great amount of researches has been conducted for IIBS and many advances have been achieved, varying illuminations makes the problem of background subtraction be in trouble. As a result, IIBS in outdoor environments is still a big challenge and it leads to a new generation of more efficient and robust surveillance systems in the coming years.

V. CONCLUSION

In this paper, a comprehensive review with a special attention to methods handling varying illuminations for outdoor surveillance has been provided. To do this, we systemically categorize existing approaches for this task into two main groups and each group is subsequently divided into several models with three different attributes. The natures and traits of each model for illumination-invariant background subtraction is explained in detail and some of methods are qualitatively and quantitatively analyzed by using two representative databases constructed for testing the performance under varying illuminations. Based on various experimental results, advantages and disadvantages of corresponding methods have been discussed with future directions for improving the performance of illumination-invariant background subtraction.

REFERENCES

[1] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.

[2] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, Jul. 1997.

[3] Y. Tian, Y. Wang, Z. Hu, and T. Huang, "Selective eigenbackground for background modeling and subtraction in crowded scenes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 11, pp. 1849–1864, Nov. 2013.

[4] A. B. Chan and N. Vasconcelos, "Counting people with low-level features and Bayesian regression," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2160–2177, Apr. 2012.

[5] K. Wang, Y. Liu, C. Guo, and F.-Y. Wang, "A multi-view learning approach to foreground detection for traffic surveillance applications," *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 4144–4158, Jun. 2016.

[6] Wahyono, A. Filonenko, and K.-H. Jo, "Unattended object identification for intelligent surveillance systems using sequence of dual background difference," *IEEE Trans. Ind. Informat.*, vol. 12, no. 6, pp. 2247–2255, Dec. 2016.

[7] T. Zhang, A. Wiliem, G. Hemson, and B. C. Lovell, "Detecting kangaroos in the wild: The first step towards automated animal surveillance," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2015, pp. 1961–1965.

[8] W. Kim and C. Kim, "Background subtraction for dynamic texture scenes using fuzzy color histograms," *IEEE Signal Process. Lett.*, vol. 19, no. 3, pp. 127–130, Mar. 2012.

[9] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," in *Proc. 7th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1. Sep. 1999, pp. 255–261.

[10] M. Piccardi, "Background subtraction techniques: A review," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, vol. 4. Oct. 2004, pp. 3099–3104.

[11] 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, Mar. 2005.

[12] Y. Benezeth, P. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Review and evaluation of commonly-implemented background subtraction," in *Proc. IAPR Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.

[13] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Comput. Vis. Image Understand.*, vol. 122, pp. 4–21, May 2014.

[14] T. Bouwmans and E. H. Zahzah, "Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance," *Comput. Vis. Image Understand.*, vol. 122, pp. 22–34, May 2014.

[15] Y. Benezeth, P. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Comparative studies of background subtraction algorithms," *J. Electron. Imag.*, vol. 19, no. 3, pp. 1–4, Jul. 2010.

[16] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking groups of people," *Comput. Vis. Image Understand.*, vol. 80, no. 1, pp. 42–56, Oct. 2000.

[17] W. Li, H. H. Yue, S. Valle-Cervantes, and S. J. Qin, "Recursive PCA for adaptive process monitoring," *J. Process Control*, vol. 10, no. 5, pp. 471–486, May 2000.

[18] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh, "Background modeling and subtraction of dynamic scenes," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 1305–1312.

[19] D.-S. Lee, "Effective Gaussian mixture learning for video background subtraction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 827–832, May 2005.

[20] E. Rivlin, M. Rudzsky, R. Goldenberg, U. Bogomolov, and S. Lepchev, "A real-time system for classification of moving objects," in *Proc. IAPR Int. Conf. Pattern Recognit.*, vol. 3. Aug. 2002, pp. 688–691.

[21] T. Parag, A. Elgammal, and A. Mittal, "A framework for feature selection for background subtraction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 1916–1923.

[22] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 1999, pp. 246–252.

[23] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, vol. 2. Aug. 2004, pp. 28–31.

[24] T. S. F. Haines and T. Xiang, "Background subtraction with Dirichlet process mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 670–683, Apr. 2014.

[25] H. Wang and D. Suter, "A consensus-based method for tracking: Modeling background scenario and foreground appearance," *Pattern Recognit.*, vol. 40, no. 3, pp. 1091–1105, Mar. 2007.

[26] M. Van Droogenbroeck and O. Paquot, "Background subtraction: Experiments and improvements for ViBe," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. Workshop*, Providence, RI, USA, Jun. 2012, pp. 32–37.

[27] M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedbacks: The pixel-based adaptive segmenter," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 38–43.

[28] P. L. St-Charles, G. A. Bilodeau, and R. Bergevin, "SuBSENSE: A universal change detection method with local adaptive sensitivity," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 359–373, Jan. 2015.

[29] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-Time Imag.*, vol. 11, no. 3, pp. 172–185, Jun. 2005.

[30] A. Doshi and M. Trivedi, "'Hybrid cone-cylinder' codebook model for foreground detection with shadow and highlight suppression," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Nov. 2006, pp. 14–19.

[31] A. Pal, G. Schaefer, and M. E. Celebi, "Robust codebook-based video background subtraction," in *Proc. IEEE ICASSP*, Mar. 2010, pp. 1146–1149.

[32] J.-M. Guo, Y.-F. Liu, C.-H. Hsia, and C.-S. Hsu, "Hierarchical method for foreground detection using codebook model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 6, pp. 804–815, Jun. 2011.

[33] J.-M. Guo, C.-H. Hsia, Y.-F. Liu, M.-H. Shih, C.-H. Chang, and J.-Y. Wu, "Fast background subtraction based on a multilayer codebook model for moving object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, pp. 1809–1821, Oct. 2013.

[34] Z. Zeng, J. Jia, Z. Zhu, and D. Yu, "Adaptive maintenance scheme for codebook-based dynamic background subtraction," *Comput. Vis. Image Understand.*, vol. 152, pp. 58–66, Nov. 2016.

[35] A. E. Waters, A. C. Sankaranarayanan, and R. Baraniuk, "SpaRCS: Recovering low-rank and sparse matrices from compressive measurements," in *Proc. Adv. Neural Inf. Process. Syst.*, Granada, Spain, Dec. 2011, pp. 1089–1097.

[36] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2011, pp. 1–9.

[37] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–39, 2011.

[38] 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. 35, no. 3, pp. 597–610, Mar. 2013.

[39] N. Wang and D.-Y. Yeung, "Bayesian robust matrix factorization for image and video processing," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, VIC, Australia, Dec. 2013, pp. 1785–1792.

[40] S. Javed, S. H. Oh, A. Sobral, T. Bouwmans, and S. K. Jung, "OR-PCA with MRF for robust foreground detection in highly dynamic backgrounds," in *Proc. Asian Conf. Comput. Vis.*, Nov. 2014, pp. 284–299.

[41] J. Yao, X. Liu, and C. Qi, "Foreground detection using low rank and structured sparsity," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2014, pp. 1–6.

[42] L. Li, P. Wang, Q. Hu, and S. Cai, "Efficient background modeling based on sparse representation and outlier iterative removal," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 2, pp. 278–289, Feb. 2016.

[43] Y. Li, J. Yan, Y. Zhou, and J. Yang, "Optimum subspace learning and error correction for tensors," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2010, pp. 790–803.

[44] W. Hu, X. Li, X. Zhang, X. Shi, S. Maybank, and Z. Zhang, "Incremental tensor subspace learning and its applications to foreground segmentation and tracking," *Int. J. Comput. Vis.*, vol. 91, no. 3, pp. 303–327, 2011.

[45] A. Sobral, S. Javed, S. K. Jung, T. Bouwmans, and E. H. Zahzah, ''Online stochastic tensor decomposition for background subtraction in multispectral video sequences,'' in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 946–953.

[46] W. Cao, Y. Wang, J. Sun, D. Meng, C. Yang, A. Cichocki, and Z. Xu, ''Total variation regularized tensor RPCA for background subtraction from compressive measurements,'' *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4075–4090, Sep. 2016.

[47] T. W. Chua, Y. Wang, and K. Leman, ''Adaptive texture-color based background subtraction for video surveillance,'' in *Proc. 19th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2012, pp. 49–52.

[48] W. Zhang, X. Z. Fang, X. K. Yang, and Q. M. J. Wu, ''Moving cast shadows detection using ratio edge,'' *IEEE Trans. Multimedia*, vol. 9, no. 6, pp. 1202–1214, Oct. 2007.

[49] M. Sedky, M. Moniri, and C. C. Chibelushi, ''Spectral-360: A physics-based technique for change detection,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Columbus, OH, USA, Jun. 2014, pp. 405–408.

[50] W. Kim and Y. Kim, ''Background subtraction using illumination-invariant structural complexity,'' *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 634–638, May 2016.

[51] A. Rafique, A. M. Sheri, and M. Jeon, ''Background scene modeling for PTZ cameras using RBM,'' in *Proc. Int. Conf. Control, Autom. Inf. Sci. (ICCAIS)*, Dec. 2014, pp. 165–169.

[52] M. Braham and M. V. Droogenbroeck, ''Deep background subtraction with scene-specific convolutional neural networks,'' in *Proc. Int. Conf. Syst. Signal Image Process.*, May 2016, pp. 1–4.

[53] R. VanRullen, ''Visual saliency and spike timing in the ventral visual pathway,'' *J. Physiol.*, vol. 97, no. 2, pp. 365–377, 2003.

[54] O. S. Packer and D. M. Dacey, ''Synergistic center-surround receptive field model of monkey H1 horizontal cells,'' *J. Vis.*, vol. 5, no. 11, pp. 1038–1054, 2005.

[55] V. Mahadevan and N. Vasconcelos, ''Spatiotemporal saliency in dynamic scenes,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 171–177, Jan. 2010.

[56] W. Kim, C. Jung, and C. Kim, ''Spatiotemporal saliency detection and its applications in static and dynamic scenes,'' *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 4, pp. 446–456, Apr. 2011.

[57] W. Kim and C. Kim, ''Spatiotemporal saliency detection using textural contrast and its applications,'' *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 4, pp. 646–659, Apr. 2014.

[58] W. Kim and J.-J. Han, ''Video saliency detection using contrast of spatiotemporal directional coherence,'' *IEEE Signal Process. Lett.*, vol. 21, no. 10, pp. 1250–1254, Oct. 2014.

[59] Z. Liu, X. Zhang, S. Luo, and O. Le Meur, ''Superpixel-based spatiotemporal saliency detection,'' *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 9, pp. 1522–1540, Sep. 2014.

[60] X. Hou and L. Zhang, ''Saliency detection: A spetral residual approach,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[61] X. Cui, Q. Liu, and D. Metaxas, ''Temporal spectral residual: Fast motion saliency detection,'' in *Proc. ACM Int. Conf. Multimedia*, Oct. 2009, pp. 617–620.

[62] C. Guo and L. Zhang, ''A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression,'' *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 185–198, Jan. 2010.

[63] J. Li, M. D. Levine, X. An, X. Xu, and H. He, ''Visual saliency based on scale-space analysis in the frequency domain,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 996–1010, Apr. 2013.

[64] 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.

[65] L. Maddalena and A. Petrosino, ''The SOBS algorithm: What are the limits?'' in *Proc. IEEE Comput. Vis. Pattern Recognit. Workshop (CVPRW)*, Jun. 2012, pp. 21–26.

[66] A. Ferone and L. Maddalena, ''Neural background subtraction for pan-tilt-zoom cameras,'' *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 44, no. 5, pp. 571–579, May 2014.

[67] B. Lei and L. Xu, ''Real-time outdoor video surveillance with robust foreground extraction and object tracking via multi-state transition management,'' *Pattern Recognit. Lett.*, vol. 27, no. 15, pp. 1816–1825, Nov. 2006.

[68] Y. Wang, K.-F. Loe, and J.-K. Wu, ''A dynamic conditional random field model for foreground and shadow segmentation,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 279–289, Feb. 2006.

[69] Z. Zivkovic and F. van der Heijden, ''Efficient adaptive density estimation per image pixel for the task of background subtraction,'' *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 773–780, 2006.

[70] B. White and M. Shah, ''Automatically tuning background subtraction parameters using particle swarm optimization,'' in *Proc. IEEE Conf. Multimedia Expo*, Jul. 2007, pp. 1826–1829.

[71] G.-A. Bilodeau, J.-P. Jodoin, and N. Saunier, ''Change detection in feature space using local binary similarity patterns,'' in *Proc. Int. Conf. Comput. Robot Vis. (CRV)*, May 2013, pp. 106–112.

[72] M. Heikkila and M. Pietikäinen, ''A texture-based method for modeling the background and detecting moving objects,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 657–662, Apr. 2006.

[73] L. Fei-Fei and P. Perona, ''A Bayesian hierarchical model for learning natural scene categories,'' in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2005, pp. 524–531.

[74] S. Lazebnik, C. Schmid, and J. Ponce, ''Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,'' in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2006, pp. 2169–2178.

[75] H. Pan, Y. Zhu, A. K. Qin, and L. Xia, ''Mining heterogeneous class-specific codebook for categorical object detection and classification,'' in *Proc. 20th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2013, pp. 3132–3136.

[76] T.-S. Chao, Y.-L. Lin, Y.-H. Kuo, and W. H. Hsu, ''Scalable object detection by filter compression with regularized sparse coding,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3900–3907.

[77] Q. Wei, X. Zhang, Y. Kong, W. Hu, and H. Ling, ''Compact visual codebook for action recognition,'' in *Proc. IEEE Conf. Image Process.*, Sep. 2010, pp. 3805–3808.

[78] Z. Wang, Y. Wang, L. Wang, and Y. Qiao, ''Codebook enhancement of vlad representation for visual recognition,'' in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2016, pp. 1258–1262.

[79] F. De la Torre and M. J. Black, ''A framework for robust subspace learning,'' *Int. J. Comput. Vis.*, vol. 54, no. 1, pp. 117–142, Aug. 2003.

[80] L. Cheng, M. Gong, D. Schuurmans, and T. Caelli, ''Real-time discriminative background subtraction,'' *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1401–1414, May 2011.

[81] F. Porikli, L. Davis, Q. Wang, Y. Li, and C. Regazzoni, ''Special issue on deep learning for video surveillance,'' *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 11, pp. 2159–2160, Nov. 2016.

[82] (Dec. 2001). *IEEE Workshop on Performance Evaluation of Tracking and Surveillance*. [Online]. Available: fpt://ftp.pets.rdg.ac.uk/ pub/pets2001

[83] J. Davis and V. Sharma, ''Background subtraction using contour based fusion of thermal and visible imagery,'' *Comput. Vis. Image Understand.*, vol. 106, no. 2, pp. 162–182, May 2007.

[84] S. Agarwal, A. Awan, and D. Roth, ''Learning to detect objects in images via a sparse, part-based representation,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1475–1490, Nov. 2004.

[85] Z. Zhou, X. Li, J. Wright, E. J. Candès, and Y. Ma, ''Stable principal component pursuit,'' in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 2010, pp. 1518–1522.

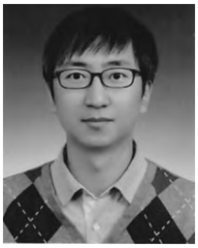
[86] S. Becker, E. J. Candès, and M. Grant, ''Templates for convex cone problems with applications to spares recovery,'' *Math. Programm. Comput.*, vol. 3, no. 3, pp. 165–218, Aug. 2011.

[87] X. Ding, L. He, and L. Carin, ''Bayesian robust principal component analysis,'' *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3419–3430, May 2011.

[88] S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos, ''Sparse Bayesian methods for low-rank matrix estimation,'' *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3964–3977, Aug. 2012.

[89] S. Bianco and R. Schettini, ''Adaptive color constancy using faces,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1505–1518, Aug. 2014.

[90] B. A. Maxwell, R. M. Friedhoff, and C. A. Smith, ''A bi-illuminant dichromatic reflection model for understanding images,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.



WONJUN KIM received the B.S. degree in electrical engineering from Sogang University, Seoul, South Korea, the M.S. degree from the Department of Information and Communications, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, and the Ph.D. degree from the Department of Electrical Engineering, KAIST, in 2006, 2008, and 2012, respectively. From 2012 to 2016, he was a Research Staff Member with the Samsung Advanced Institute of Technology, Gyeonggi-do, South Korea. Since 2016, he has been with the Department of Electronics Engineering, Konkuk University, Seoul, where he is currently an Assistant Professor. His research interests include image and video understanding, computer vision, pattern recognition, and biometrics, with an emphasis on background subtraction, saliency detection, and face and action recognition. He has served as a Regular Reviewer for over 30 international journal papers, including the IEEE Transactions on Image Processing, the IEEE Transactions on Circuits and Systems for Video Technology, the IEEE Transactions on Multimedia, IEEE Transactions on Cybernetics, the IEEE Signal Processing Letters, and Pattern Recognition.

CHANHO JUNG received the B.S. and M.S. degrees in electronic engineering from Sogang University, Seoul, South Korea, in 2004 and 2006, respectively, and the Ph.D. degree in electrical engineering from the Korea Advanced Institute of Science Technology, Daejeon, South Korea, in 2013. From 2006 to 2008, he was a Research Engineer with the Digital Television Research Laboratory, LG Electronics, Seoul. From 2013 to 2016, he was a Senior Researcher with the Electronics and Telecommunications Research Institute, Daejeon. Since 2016, he has been with the Department of Electrical Engineering, Hanbat National University, Daejeon, where he is currently an Assistant Professor. His current research interests are computer vision, machine learning, embedded system, pattern recognition, and image processing.