IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 27, 2017, accepted November 14, 2017, date of publication November 17, 2017,
date of current version December 22, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2775040

Thermal Sensor-Based Multiple Object Tracking for Intelligent Livestock Breeding

WONJUN KIM[1], (Member, IEEE), YONG BEOM CHO[1], (Member, IEEE), AND SANGRAK LEE[2]
[1]Department of Electronics Engineering, Konkuk University, Seoul 05029, South Korea
[2]Department of Animal Science and Technology, Konkuk University, Seoul 05029, South Korea

Corresponding author: Yong Beom Cho (ybcho@konkuk.ac.kr)

This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries through the Advanced Production Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs under Grant 116056-03.

ABSTRACT Visual object tracking is an essential technique for constructing intelligent livestock management systems. Behavior patterns estimated from the trajectories of animals provide substantial useful information related to estrus cycle, disease prognosis and so on. However, similar colors and shapes between animals often lead to the failure of tracking multiple objects, and the background clutter of the breeding space further makes the problem intractable. In this paper, we propose a novel method for tracking animals using a single thermal sensor. The key idea of the proposed method is to represent the foreground (i.e., animals) easily obtained by a simple thresholding in a thermal frame as a topographic surface, which is very helpful for finding the boundary of each object even in cases with overlapping. Based on the segmentation results derived from morphological operations on the topographic surface, the center positions of all the animals are consistently updated with an efficient refinement scheme that is robust to the abrupt motions of animals. Experimental results using various thermal video sequences demonstrate the efficiency and robustness of our method for tracking animals in a breeding space compared to previous approaches proposed in the literature.

INDEX TERMS Visual object tracking, intelligent livestock management, thermal sensor, topographic surface-based segmentation, overlapped cases.

I. INTRODUCTION

With the rapid increase in the use of various sensors for monitoring diverse environments, video surveillance techniques have been actively applied to a wide range of real-world scenarios, e.g., safety in public spaces [1], crowd analysis [2], and traffic control [3], over the last few decades. Thanks to recently developed operations, such as object detection and recognition under dynamic outdoor environments (e.g., illumination changes and moving backgrounds), video surveillance techniques are now ready to be applied to livestock management systems, systems that are totally different from conventional systems that mainly observe the motion patterns of people. For example, the number of times animals drink water or the amount of movement need to be automatically analyzed by such a system without any human interaction. Performing visual object tracking is a key technique for the success of such intelligent livestock breeding systems. In contrast to general tracking scenarios, animals in a breeding space, e.g., cows and pigs, have similar appearances (i.e., colors and shapes), and furthermore, they often remain in close proximity, thereby yielding significant occlusions, which often lead to tracking failures. To resolve this problem, most previous approaches have utilized radio-frequency identification (RFID) sensors, which can be easily attached to the body parts of animals (e.g., ear or neck) [4]–[6]. However, the operating range of RFID sensors is very limited, and thus, the movements of animals irregularly occurring in the whole breeding space cannot be completely covered. In the field of computer vision, many researchers have focused on precisely analyzing motion patterns under complex environments based on robust optical flow [7]–[9]. Specifically, they adopted various optimization techniques to reduce the ambiguity of motions, which can be help for segmenting behavior patterns of animals. On the other hand, some studies attempted to integrate such optical flow algorithms into the hardware for the system implementation [10]–[12].

In this paper, we propose a novel method for multiple-object tracking based on topographic surface analysis. One important advantage of using a thermal sensor instead of a

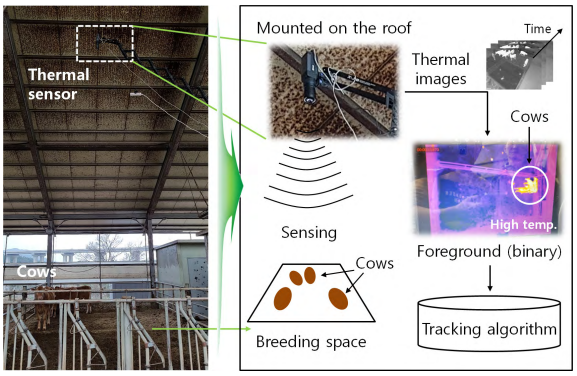


FIGURE 1. Our intelligent livestock (e.g., cows) management system based on the thermal sensor. Notice that the thermal sensor is installed on the roof, which is vertical to the ground for monitoring the movements of all the animals in the whole breeding space.

traditional RGB sensor is that the foreground (i.e., animals) can be efficiently extracted from the background by a simple thresholding even at night since animals have a relatively high body temperature compared to the ground regardless of lighting conditions. To the best of our knowledge, this is the first work to apply thermal object tracking to livestock management systems. Specifically, we first installed a thermal sensor in the vertical direction from the ground for monitoring the whole breeding space, as shown in Fig. 1. In the following, we convert extracted foreground regions whose shapes are almost convex (see Fig. 2) into a topographic surface based on a distance transform [13]. It is noteworthy that our scheme has a great ability to separate overlapped animals since such overlaps form concave parts at the boundary between different objects (see Fig. 4), which are well revealed on the topographic surface. Finally, the center position of each separated region is consistently updated with a simple refinement scheme particularly devised to handle the abrupt motions of animals. The main contributions of this paper are threefold:

- In contrast to traditional surveillance systems mostly based on RGB camera sensors, we construct a thermal-sensor-based livestock management system, which allows the proposed tracking algorithm to work even at night. This is fairly desirable for correctly analyzing the behavior patterns of livestock regardless of the outdoor environment.
- Based on the observation that the shape of an animal is almost convex from a top-down view, we propose to exploit our tracking algorithm using topographic surface analysis, which successfully reveals concave points generated by overlaps between animals (such points are actually associated with the boundary). Therefore, the proposed method significantly improves the tracking performance even when addressing the complex motion patterns of animals, whereas previous approaches suffer from such overlap problems.
- We estimate the center position of each foreground region (i.e., animal) with pixels inside the boundaries precisely obtained from the topographic surface-based segmentation rather than by exploiting the bounding box generally employed in previous tracking methods. Furthermore, the corresponding positions are consistently updated with a simple refinement scheme, which helps to overcome issues caused by the abrupt motions of animals.

The remainder of this paper is organized as follows. A comprehensive review of previous tracking methods is presented in Section II. The proposed tracking method based on topographic surface analysis is introduced in detail in Section III. We evaluate the performance of the proposed method with various thermal video sequences and compare our method with other approaches in Section IV. The conclusions follow in Section V.

II. RELATED WORK

In this section, we give a brief review of visual object tracking methods introduced in the field of computer vision. Since visual object tracking is a key prerequisite for further applications, such as object recognition and behavior understanding, many researchers have devoted considerable efforts to constructing reliable tracking methods.

Initially, researchers focused on explicitly estimating the statuses of target objects (i.e., position and shape) based on Kalman and particle filtering techniques [14]–[18]. However, most such methods are vulnerable to slight changes in the target objects since they employ static appearance models. On the other hand, some studies have focused on estimating the density distribution of features in a target region [19], [20] to track the region consistently. Although these models are robust to variations in the target object, they are weak to feature ambiguities driven by deformations and occlusions between animals. To resolve this problem, adaptive models, which evolve during the tracking process by utilizing the target region and its surroundings, have been introduced [21], [22]. As representative methods among a vast number of studies, Grabner *et al.* [23] proposed to define tracking as an online classification problem, which has the distinct advantage of being able to be adaptive to appearance changes in the target object. To do this, they developed a novel online AdaBoost algorithm with fast computable features (e.g., Haar-like wavelets, orientation histograms, and local binary patterns). In a similar spirit, Babenko *et al.* [24] also trained a discriminative classifier in an online manner to separate objects from the background. Furthermore, to reduce the drift problem, they proposed to exploit multiple instance learning instead of traditional supervised learning models. Hare *et al.* [25] attempted to avoid the intermediate classification step, which does not actually coincide with the objective of the tracker, based on a structured output prediction scheme. For the definition of the prediction model, they adopted a kernelized structured output support vector machine (SVM), which was also built in an online manner. Kalal *et al.* [26] proposed to combine the tracking procedure with learning and detection, which makes the tracking

performance robust to drift problems and variations in the target object. To this end, they guided a sampling process of the boosting classifier using structural constraints. Most notably, Henriques *et al.* [27] proposed to exploit the kernelized correlation filters to achieve fast computation based on their observation that samples obtained from the neighbor regions of the target object can be represented as a circulant matrix. Note that they efficiently reduced the processing time as well as storage requirements by diagonalizing this matrix with the fast Fourier transform (FFT). On the other hand, some studies have developed robust optical flow algorithms for segmenting multiple motions. Specifically, Heas *et al.* [7] proposed to select the optimal model and corresponding hyper-parameters for estimating motions in a generic Bayesian framework. Mohamed *et al.* [8] proposed to adopt the texture descriptor for modeling an illumination-robust constancy rather than the brightness constancy, leading to the significant increase of the accuracy for constructing the motion field. Crivelli *et al.* [9] proposed a novel inverse integration scheme, which shows the robustness to input noise and better stability properties compared to the traditional Euler integration method. Although these methods have achieved remarkable improvements for object tracking, they still suffer from issues arising from long-time occlusions and background clutter.

To efficiently overcome such problems, several researchers have begun to employ thermal sensors alongside traditional RGB cameras. Specifically, Leykin *et al.* [28] defined a background model based on a multi-modal distribution of colors and temperatures, which can be efficiently incorporated into a particle filter maximizing the probability of the scene model with a number of samples obtained from the model probability space. Talha and Stolkin [29] also proposed a particle-filter-based tracking scheme for camouflaged targets by adaptively combining data from thermal and visible spectra cameras. Li *et al.* [30] conducted the multi-task joint sparse coding for combining grayscale and thermal modalities. They further incorporated the reliability constraint into the Laplacian sparse representation for adaptively fusing different types of source data. Although they attempted to alleviate the classic problems of visual object tracking, most methods still require color information, which is not available at night. On the other hand, some studies have focused on developing single thermal sensor-based tracking algorithms. Zhang *et al.* [31] exploited the edge information obtained from the target object in a partial manner to efficiently describe the shape without colors in the thermal image. Gundogdu *et al.* [32] attempted to apply deep convolutional neural network (CNN) features to the correlation filter, which are trained on the thermal image sequences, for improving the tracking performance. Berg *et al.* [33] proposed a template-based scheme with the background update to avoid contamination of background in terms of the object template in the thermal video sequence. Kwak *et al.* [34] employed the online learning strategy using the boosted random fern, which is constructed based on the temperature difference in the thermal image. In this paper, we adopt a single thermal sensor for this task. Rather than tracking people or cars, as is well studied in previous approaches, we focus on the motion patterns of animals obtained from the thermal sensor for constructing the intelligent livestock breeding system. Technical details will be introduced in the following section.

III. PROPOSED METHOD

The motivation of our new approach is that convex-like objects (e.g., animals captured from a top-down viewpoint) tend to form concave points when they are overlapped or occluded, which can be efficiently detected based on the topographic surface. It follows that a measure of the distance from the center position of each object provides a good approximation for finding the segmentation line related to concave points in a single connected component. Based on pixels belonging to the same segmented region, the center position of each object can be correctly computed and consistently updated with the simple refinement scheme even in severely overlapped cases as well as considering the abrupt motions of animals.

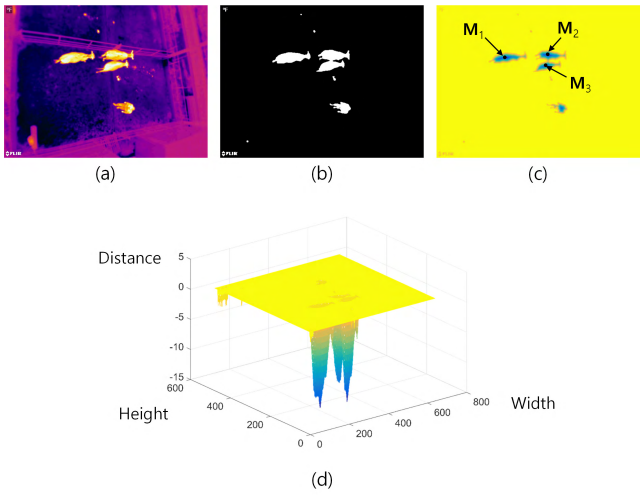


FIGURE 2. (a) Original input frame from the thermal sensor installed in the cow breeding space. (b) Binarized result obtained by a simple thresholding. (c) Result of distance transform. (d) Topographic surfaces.

A. TOPOGRAPHIC SURFACE-BASED OBJECT SEGMENTATION

Due to the power of the thermal sensor, we can easily obtain the binarized image $B(x, y)$ of each frame by a simple thresholding on the original input $I(x, y)$, i.e., $B(x, y) = 1$ if $I(x, y) > \tau$ (e.g., $\tau = 150$) and $B(x, y) = 0$ otherwise. In the following, we define our topographic surface by applying the distance transform to the binarized image, as shown in Fig. 2. It is noteworthy that the derived topographic surface can be regarded as multiple watersheds (see Fig. 2(d)).

Motivated by several studies employing the concept of a topographic surface to separate overlapped cells in medical images [35], [36], we propose to adopt the marker-controlled watershed segmentation technique [37] for our tracking

purpose. More specifically, let $\mathbf{M}_1, \mathbf{M}_2, \cdots, \mathbf{M}_R$ denote the positions of the regional minima in the watershed, which are given by the center position of each object (i.e., marker), in the distance image $D(x, y)$ (see Fig. 2(c)) where R is the total number of target objects. Notice that the distance transform yields the smallest value at the center of each connected component in our implementation. $C(\mathbf{M}_i)$ denotes a set of pixels belonging to the same watershed associated with regional minimum $\mathbf{M}_i$, where i is the index for each watershed (i.e., object). Notice that all the pixels in the catchment basin form a connected component. Finally, we let $S[\epsilon]$ denote a set of coordinates (p, q) satisfying the following conditions [38]:

$$S[\epsilon] = \{(p, q) | D(p, q) < \epsilon\}. \quad (1)$$

Here, $S[\epsilon]$ contains all the pixels lying below the plane $D(x, y) = \epsilon$ in a geometrical view. Conceptually, ϵ and $S[\epsilon]$ can be regarded as the water level and the surface of the catchment basin, respectively. Based on this, we define a submerged region $C_\epsilon(\mathbf{M}_i)$ as a set of pixels (x, y) when the water level goes up to ϵ, given by

$$C_\epsilon(\mathbf{M}_i) = \{(x, y) | (x, y) \in C(\mathbf{M}_i) \text{ and } (x, y) \in S[\epsilon]\}. \quad (2)$$

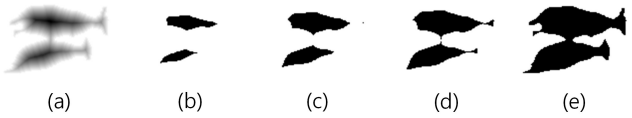


FIGURE 3. Change in the catchment basin according to the water level. (a) Topographic surface (i.e., distance map). Notice that the pure black denotes the smallest distance. (b) $\epsilon = 1$. (c) $\epsilon = 5$. (d) $\epsilon = 7$. (e) $\epsilon = 10$.

Subsequently, we also define the union of catchment basins as $C[\epsilon] = \bigcup_{i=1}^{R} C_\epsilon(\mathbf{M}_i)$. In this work, we set ϵ based on the value onto each marker position (m, n), e.g., $\epsilon = 1.3 \times D(m, n)$. Some examples of $C_\epsilon(\mathbf{M}_i)$ maps with different water levels ϵ are shown in Fig. 3. Notice that pixels belonging to $C_\epsilon(\mathbf{M}_i)$ are indicated in black. As the water rises gradually from the river bed (i.e., the marker position), nearby catchment basins may become merged at a specific level, and thus, a dam, i.e., a boundary between different watersheds, needs to be constructed to prevent overflow. To build a one-pixel-thick dam, we employ the morphological approach [38]. Briefly, the dilating operation is iteratively applied to $C_\epsilon(\mathbf{M}_i)$ belonging to the same connected component in $S[\epsilon]$ until the expanded regions from different watersheds (i.e., objects) are overlapped, as shown in Fig. 4. It is worth noting that this overlap occurs at the concave point (see the arrows in Fig. 4(d)). Such overlapped regions are determined as the boundary between nearby objects and can be formulated as follows:

$$L(x, y) = \begin{cases} 1, & \text{if } (x, y) \in \{\tilde{C}_\epsilon^t(\mathbf{M}_i) \cap \tilde{C}_\epsilon^t(\mathbf{M}_j)\} \text{ or} \\ & (x, y) \in C^B[\epsilon], \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

FIGURE 4. (a) Surfaces of two catchment basins (pink regions), e.g., $C_\epsilon(\mathbf{M}_1)$ and $C_\epsilon(\mathbf{M}_2)$. Notice that the black region is one of the connected components belonging to $S[\epsilon]$. (b)(c) Intermediate results of dilation using $C_\epsilon(\mathbf{M}_1)$ and $C_\epsilon(\mathbf{M}_2)$. (d) Results of building a dam (i.e., watershed line), which is also defined at the concave point (see arrows). For a better view, we assign different colors to each object, while the dam is represented by red-colored lines.

where $\tilde{C}_\epsilon^k(\mathbf{M}_i)$ denotes the t times dilated watershed related to the i^{th} marker and $C^B[\epsilon]$ indicates the boundary pixels of submerged regions defined by ϵ. Therefore, we can compute the new center position of each object at the next frame as follows:

$$\mathbf{M}_i(k+1) = \frac{1}{Card(\tilde{C}_\epsilon^S(\mathbf{M}_i(k)))} \sum_{\mathbf{q} \in \tilde{C}_\epsilon^S(\mathbf{M}_i(k))} \mathbf{q}, \quad (4)$$

where S indicates the status of the catchment basin when the dam is constructed and k denotes the frame index, respectively. $\mathbf{q}$ is the pixel position belonging to the corresponding catchment basin. As shown in Fig. 4(d), our watershed-based scheme is capable of precisely extracting target objects with tracked markers $\mathbf{M}_i$. In the following subsection, the method for refining estimated positions will be explained in detail.

B. VELOCITY WEIGHT AND THE RIVER-BED REFINEMENT

Based on the proposed topographical segmentation, target objects can be robustly tracked in most cases. However, the motion of livestock is often nonlinear in the breeding space, e.g., movement speed is getting faster, which leads to the tracking failure. To compensate such nonlinearity, we propose to apply the velocity weight to the estimated position by (4) as follows:

$$\tilde{\mathbf{M}}_i(k+1) = \mathbf{M}_i(k+1) + \lambda \mathbf{v}_i(k+1), \quad (5)$$

where k denotes the frame index as mentioned and $\mathbf{v}_i(k)$ is the velocity of the i^{th} watershed. λ is a weight for the velocity, which is set to 0.5 in our implementation. Notice that the initial positions of the markers are manually given at the starting frame of the tracking (i.e., $\mathbf{M}_i(0)$), and the center position of each object acts as a marker in subsequent frames. Therefore, the velocity $\mathbf{v}_i(k)$ can be formulated as follows:

$$\mathbf{v}_i(k+1) = \mathbf{M}_i(k+1) - \mathbf{M}_i(k). \quad (6)$$

Notice that the effect of our velocity weight will be explained in the following subsection. Since the velocity-weighted marker position may not be perfectly matched with the valley points in our topographic surface due to the abrupt motions of animals, we propose to simply refine our result based on the spatial proximity with valley points in the distance map (i.e., topographic surface). Specifically, we find the spatially nearest point in the valley (e.g., $C_5(\mathbf{M}_i)$) of the topographic

surface with the estimated position in (5) and simply update our result as follows:

$$\tilde{\mathbf{M}}_i(k+1) = \underset{\tilde{\mathbf{b}}}{\operatorname{argmin}} ||\tilde{\mathbf{b}} - \mathbf{M}_i(k+1)||, \tag{7}$$

where $\tilde{\mathbf{b}}$ denotes the point belonging to the valley of the topographic surface. Our watershed-based segmentation, which was explained in the previous subsection, is subsequently conducted based on the corresponding position (i.e., $\tilde{\mathbf{b}}$ minimizing the right part of (7)) in subsequent frames.

For the sake of completeness, the overall procedure of the proposed method is summarized in Algorithm 1.

Algorithm 1 Topographic-surface-based object tracking in thermal videos

Data: $\mathbf{M}_i(0)$: initial center positions (i.e., river bed)
k: index of frames, R: number of animals
Result: Trajectories of multiple objects
while *until the video terminates* **do**
  1. Binary image generation
  i) Generate $B(x, y)$ with a predefined threshold
  while $i \leq R$ **do**
    2. Topographic surface-based object segmentation
    i) Compute the distance map $D(x, y)$
    ii) Set catchment basins with markers $C_\epsilon(\mathbf{M}_i)$ where $i \in \{1, 2, \cdots, R\}$
    iii) Conduct the morphological dilation to construct a dam (i.e., segmentation)

$$L(x,y) = \begin{cases} 1, (x,y) \in \{\tilde{C}^t_\epsilon(\mathbf{M}_i) \cap \tilde{C}^t_\epsilon(\mathbf{M}_j)\} \\ \quad \text{or } (x,y) \in C^B[\epsilon] \\ 0, \text{otherwise} \end{cases}$$

    iv) Compute the new center position using (4)
    3. Marker refinement
    i) Compute new positions of makers via velocity
    $\tilde{\mathbf{M}}_i(k+1) = \mathbf{M}_i(k+1) + \lambda \mathbf{v}_i(k+1)$ (see (5))
    where $\mathbf{v}_i(k+1) = \mathbf{M}_i(k+1) - \mathbf{M}_i(k)$
    ii) Refine estimated makers
    $\tilde{\mathbf{M}}_i(k+1) = \underset{\tilde{\mathbf{b}}}{\operatorname{argmin}} ||\tilde{\mathbf{b}} - \mathbf{M}_i(k+1)||$
  $i = i + 1$ (object index)
  end
  $k = k + 1$ (frame index)
end

C. IMPLEMENTATION DETAILS

In this subsection, we provide the process of setting parameters and implementing the proposed tracking method in detail. Specifically, we have three main parameters, i.e., a threshold value for binarization τ, an initial water level ϵ, and a weight for the velocity λ. For binarization, a widely-used adaptive thresholding scheme [39] is employed, thus τ is automatically set for every frame. In the procedure of segmentation, first of all, to extract the area

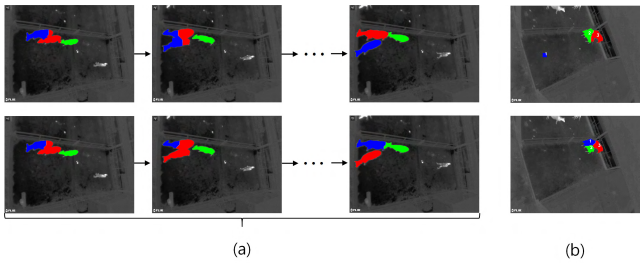


FIGURE 5. Effect of the velocity weight. (a) Results without the velocity weight (top) and with the velocity weight $\lambda = 0.5$ (bottom). (b) Result by the velocity weight $\lambda = 1$ (top) and $\lambda = 0.5$ (bottom). Colors denote the identities of each cow. Notice that identities between red and blue are exchanged without the proposed velocity weight.

which we are sure such regions belong to objects (i.e., catchment basins), we conduct the distance transform based on binarization results and apply a proper threshold as the initial water level (see (1) and (2)). According to the general scheme of watershed [40], [41], ϵ is set to $1.3 \times D_{\min}$ where $D_{\min}$ denotes the smallest distance in each object region. To cover the accelated motion of animals, we attempt to apply the velocity weight to the newly estimated position. As shown in Fig. 5(a), identities between nearby cows can be exchanged due to accelerated motions. In our implementation, λ is set to 0.5 throughout many experiments. If the velocity-weighted position is not out of foreground regions, the proposed method converges (i.e., yields segmentation results). Notice that too large weight (e.g., $\lambda > 0.5$) also leads to the tracking failure in some cases due to the over-accelerated speed (see Fig. 5(b)), which drives the velocity-weighted position towards background.

To implement the proposed method, the float-typed data is sufficient to allow for the sub-pixel accuray in tracking since most operations are related to estimating the position of pixels onto the discrete grid space. The overall steps of implementation can be summarized as follows: initial positions of each cow to be tracked are given at the beginning of the test video. Based on given positions, segmentation is conducted using the proposed topographical approach explained in previous subsections. Then, we update center positions of each cow based on newly segmented results and such positions are refined with the velocity weight and our river-bed refinement scheme. In the following frames, those procedures are repeated while maintaining the identity of each cow throughout the whole video sequence. The complexity of the proposed method is highly dependent on the distance transform and iterative dilations used in the topographical segmentation, thus the burden of such operations increase in proportion to the size and the number of objects to be tracked. It is noteworthy that the complexity of the proposed method can be efficiently reduced with the multi-threaded image processing (e.g., GPU-based acceleration) since the proposed algorithm can be applied to each object independently. In regard to the calculation precision, the float-type data is sufficient since the main operations of the proposed

method are mostly related to morphological dilations and pixel-wise distance computations as mentioned.

IV. EXPERIMENTAL RESULTS

In this section, we demonstrate the efficiency and robustness of the proposed method for tracking animals in the breeding space based on the thermal sensor. As the first step, we analyzed the motion patterns of cows, and this can be directly applied to other animals (e.g., pigs and horses) without any significant changes. To this end, we first installed the thermal sensor on the roof of the breeding space to reduce the projection effect, as shown in Fig. 1. For this experiment, a single thermal sensor, manufactured by FLIR (FLIR A615), is employed. Notice that this sensor is capable of capturing a given scene at VGA resolution (i.e., 640 × 480 pixels) with a high thermal sensitivity (<50 mK). The size of the experimental breeding space in the livestock house shown in Fig. 6 is 5 m×10 m, and motion patterns of three cows in this space are obtained for this experiment.

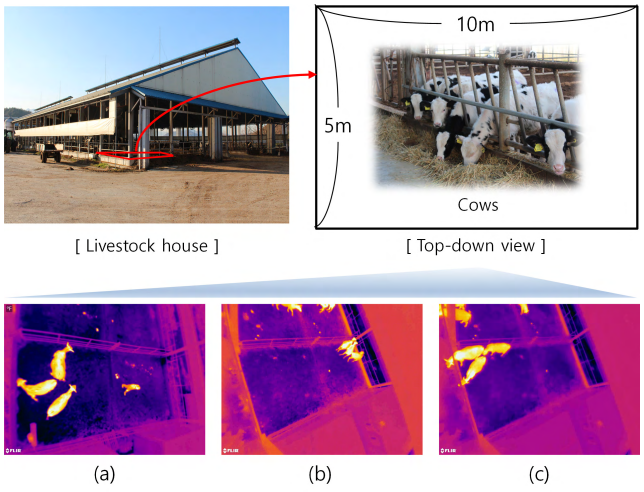


FIGURE 6. Some examples from our test sequences (640 × 480 pixels). Notice that we apply the proposed tracking method to three cows in the breeding space, whose size is 5 m×10 m, based on thermal images captured from a top-down viewpoint. (a) Seq. 1. (b) Seq. 2. (c) Seq. 3. (d) Seq. 4.

A. QUALITATIVE EVALUATION

First, we evaluate the performance of the proposed method using our three test sequences, which were captured on different days. Those video sequences are composed of 630 frames (Seq. 1), 1,000 frames (Seq. 2), and 1,000 frames (Seq. 3). Notice that Seqs. 2 and 3 are captured from slightly higher locations compared to Seq. 1 to cover a wide range of the breeding space. Some examples of our test sequences are shown in Fig. 6. To provide a better comparison, we indicate our results using the bounding box (similar to previous approaches) defined by the center position of each segmented region. Notice that the complete results obtained by the proposed topographic surface segmentation are shown in Fig. 10. To show the robustness of the proposed method, we compared our method with other tracking approaches popularly employed in the field of computer vision, i.e., BOOSTING [23], MIL [24], MEDIAN [42], TLD [26], and KCF [27], as shown in Fig. 7. It is worth noting that some of these models have often been applied to thermal infrared video sequences [43]. As shown in Fig. 7, previous methods often fail to track cows due to overlaps and the fast motions of animals. Specifically, when three cows move together in the corner space (see the third column of Fig. 7), most of the previous methods loss the identity of each cow due to the severe occlusion. In the case of TLD [26] (see Fig. 7(e)), tracking loss occurs even at the beginning part of the test sequence since the detection module of this method becomes confused due to the similar intensity of each object. Although KCF [27] yields reliable results among the previous approaches, it suffers from issues caused by fast motions, as shown in the sixth image of Fig. 7(f). It should be emphasized that the proposed method successfully estimates the center position of each object, as shown in Fig. 7(g), and furthermore provides the boundary of each object accurately based on our topographic surface analysis (see Fig. 10); in contrast, other schemes are vulnerable to the deformation of non-rigid objects. We also demonstrate the tracking results of Seqs. 2 and 3 by KCF [27] and those of the proposed method in Fig. 8. When the cows remain relatively static, KCF reliably tracks multiple objects, as shown in the results for Seq. 3 (see the right four images in Fig. 8(b)); however, it still misses target objects under abrupt changes in motion patterns (see the results for Seq. 2). In contrast, the proposed method is able to address such nonlinear motions based on our marker refinement scheme. Moreover, the performance comparison under conditions of severe overlap is also shown in Fig. 9. In this case, most of the previous methods cannot address the ambiguity from merge and split actions, while the proposed method successfully retains the identities of the cows. Therefore, it is thought that our tracking strategy can be applied to various livestock breeding systems.

B. QUANTITATIVE EVALUATION

For the quantitative analysis with the tracking algorithms employed in the previous subsection, we adopted three metrics of the global multiple-object tracking accuracy (GMOTA) [44], [45], i.e., false negative (FN), false positive (FP), and global identity miss match (GMME). These metrics are defined as follows: $\mathrm{FN} = \sum_n \frac{m_n}{g_n}$, $\mathrm{FP} = \sum_n \frac{fp_n}{g_n}$, and $\mathrm{GMME} = \sum_n \frac{ids_n}{g_n}$, where g_n denotes the number of ground truth cows at the n^{th} frame, m_n is the number of lost cows, fp_n is the number of falsely tracked cows, and ids_n is the number of identity switches in a given frame. Based on these metrics, we compared our method with previous methods quantitatively, as shown in Table 1. Notice that lower values of the three metrics are desirable for robust object tracking. Clearly, the proposed method reliably provides the trajectories of multiple cows, whereas the other approaches often track incorrect targets due to the similar intensities

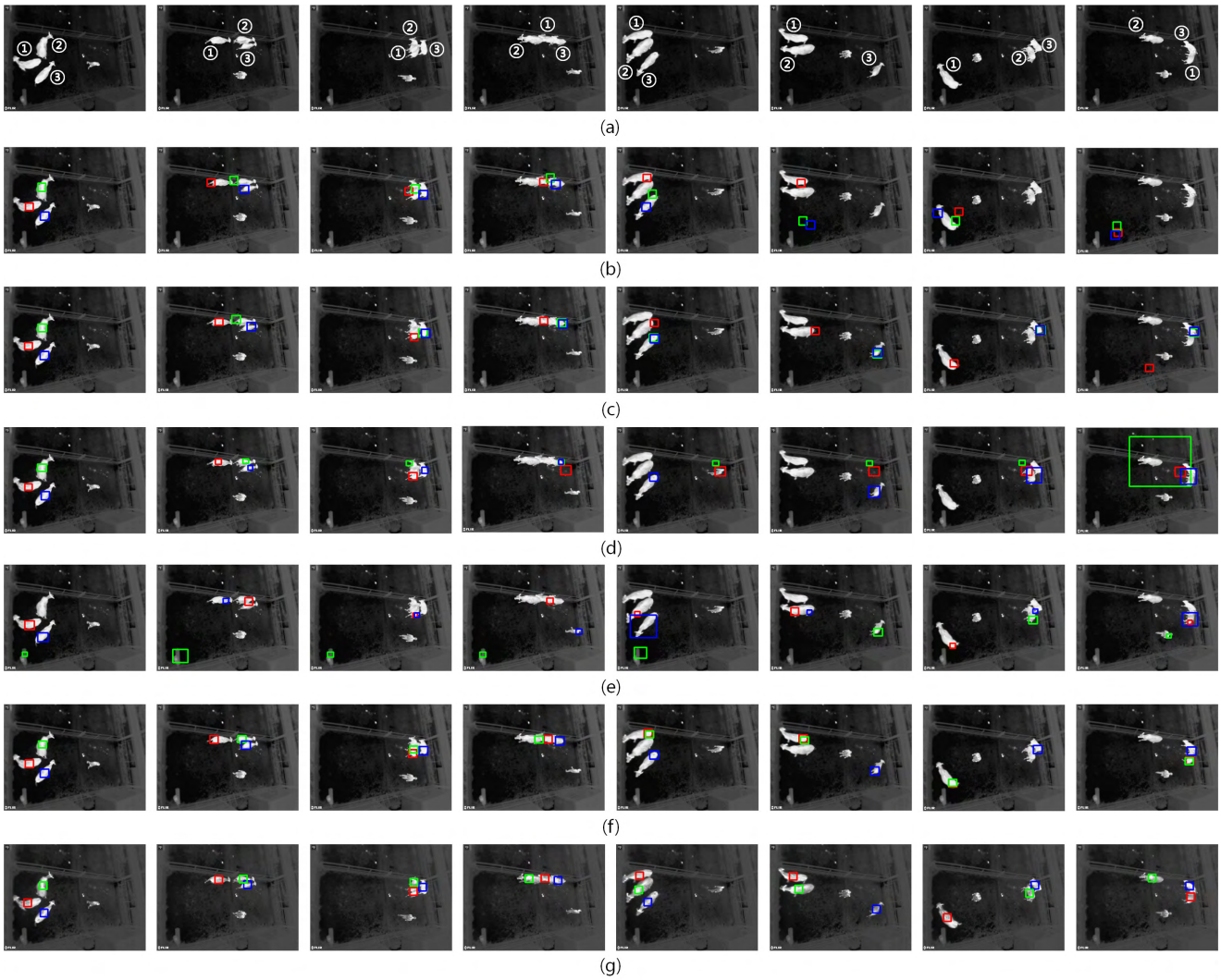


FIGURE 7. Comparison of the tracking results obtained from Seq. 1. (a) Ground truth (①, ②, and ③ are indicated in red, green, and blue, respectively). (b) BOOSTING [23]. (c) MIL [24]. (d) MEDIAN [42]. (e) TLD [26]. (f) KCF [27]. (g) Proposed method. Notice that the proposed method successfully tracks three cows even under merged cases, whereas the other approaches fail to estimate the position of each cow. Red, green, and blue denote the identity of each cow. Best viewed in color.

obtained from the thermal sensor, which leads to the high false positive rates shown in Table 1. Moreover, we checked the completeness of all the tracking methods, which indicates how completely the ground truth trajectories are tracked by a given algorithm, with metrics defined as follows [46]:

- MT : percentage of ground truth trajectories that are covered by the tracker across more than 80% of the frames.
- ML : percentage of ground truth trajectories that are covered by the tracker across less than 20% of the frames.
- PT : 1 − MT − ML.

The performance comparison based on the completeness metrics is also as shown in Table 2. In addition, the tracking accuracy based on the number of correctly tracked cows is shown in Fig. 11. As can be seen, our approach successfully tracks actively moving cows without significant losses, whereas most previous tracking models fail to estimate the positions of cows under complex motion patterns, particularly those contained in test sequence 1. Based on the various evaluation results, we confirm that the proposed tracking method is very suitable for monitoring the motion patterns of animals in the breeding space. By exploiting such motion trajectories, we can efficiently analyze the behavior patterns of livestock for disease prevention, estrus detection, etc.

TABLE 1. Performance comparison based on our test sequences.

Methods	FN ↓	FP ↓	GMME ↓
BOOSTING [23]	0.651	0.743	0.161
MIL [24]	0.721	0.595	0.173
MEDIAN [42]	0.677	0.005	0.065
TLD [26]	0.804	0.784	0.315
KCF [27]	0	0.275	0.131
Proposed method	0	0.059	0.029

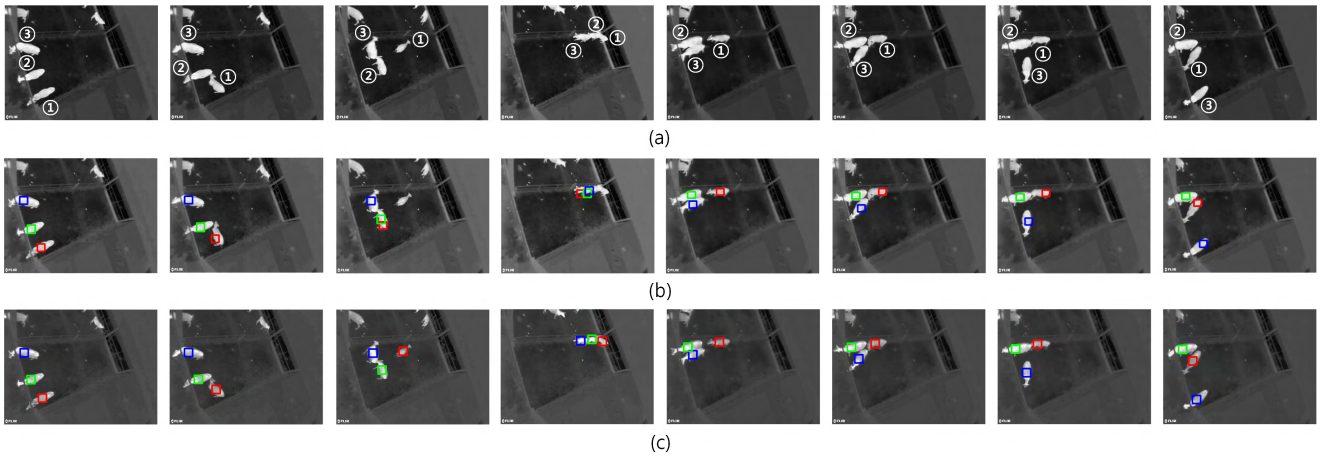


FIGURE 8. Comparison of tracking results obtained from Seqs. 2 (left four images) and 3 (right four images). (a) Ground truth. (b) KCF [27]. (c) Proposed method. The fast motions frequently generated by cows lead to tracking failures under the KCF [27], whereas such motions can be well handled by the proposed method. Notice that ①, ②, and ③ are indicated in red, green, and blue, respectively. Best viewed in color.

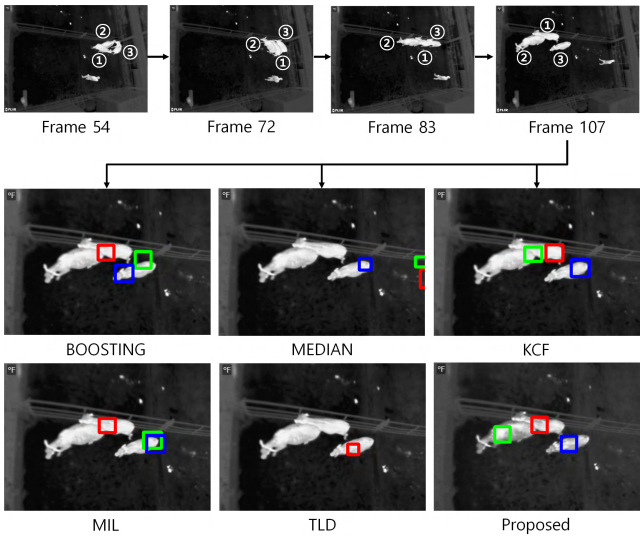


FIGURE 9. Tracking results under merge and split actions between cows. Notice that ①, ②, and ③ are indicated in red, green, and blue, respectively. Best viewed in color.

TABLE 2. Performance comparison based on completeness metrics.

Methods	MT ↑	ML ↓	PT
BOOSTING [23]	0.444	0.111	0.0.445
MIL [24]	0.667	0.111	0.222
MEDIAN [42]	0.556	0.222	0.222
TLD [26]	0.333	0.556	0.111
KCF [27]	0.778	0	0.222
Proposed method	0.889	0	0.111

To show the robustness of the proposed method in more detail, we have tested our algorithm with more stimuli. Specifically, the experiment of changing contrast, which is defined as $(E_1 - E_2)/(E_1 + E_2)$ where E_1 and E_2 denote the brightest and the darkest level of the stimulus [10], has been conducted and the tracking accuracy is evaluated accordingly. For computing this accuracy, we have manually labeled the center position of each frame for the test sequence 1 (i.e., ground truth) and computed the distance between estimated center positions and ground truth. Corresponding results are shown in Fig. 12. From the result, the proposed method works reliably even under the significant change of the contrast whereas the previous method (e.g., KCF [27]) fails to yield the tracking performance consistently. Furthermore, the errors of the proposed method is analyzed in terms of the flow metric [47], which is formulated as $s = \cos^{-1}(\mathbf{v}_e \cdot \mathbf{v}_g)$ where $\mathbf{v}_e$ and $\mathbf{v}_g$ denote the trajectory vectors of estimated center positions and ground truth, respectively. Notice that such vectors are normalized with their magnitude. The performance comparison based on the flow metric is shown in Table 3.

TABLE 3. Performance comparison in terms of the flow metric.

Method	Mean Barron error
BOOSTING [23]	5.09°
MIL [24]	5.39°
MEDIAN [42]	6.07°
TLD [26]	5.85°
KCF [27]	4.89°
Proposed method	4.05°

TABLE 4. Performance comparison according to the processing speed.

Method	Processing speed	Implementation
BOOSTING [23]	15.92 fps	C++
MIL [24]	3.38 fps	C++
MEDIAN [42]	40.37 fps	C++
TLD [26]	1.86 fps	C++
KCF [27]	49.53 fps	C++
Proposed method	32.02 fps	C++

The framework of the proposed method was implemented on a single PC (Intel i7 2.5 GHz CPU and 8 GB of RAM without parallel processing) with Visual Studio 2015

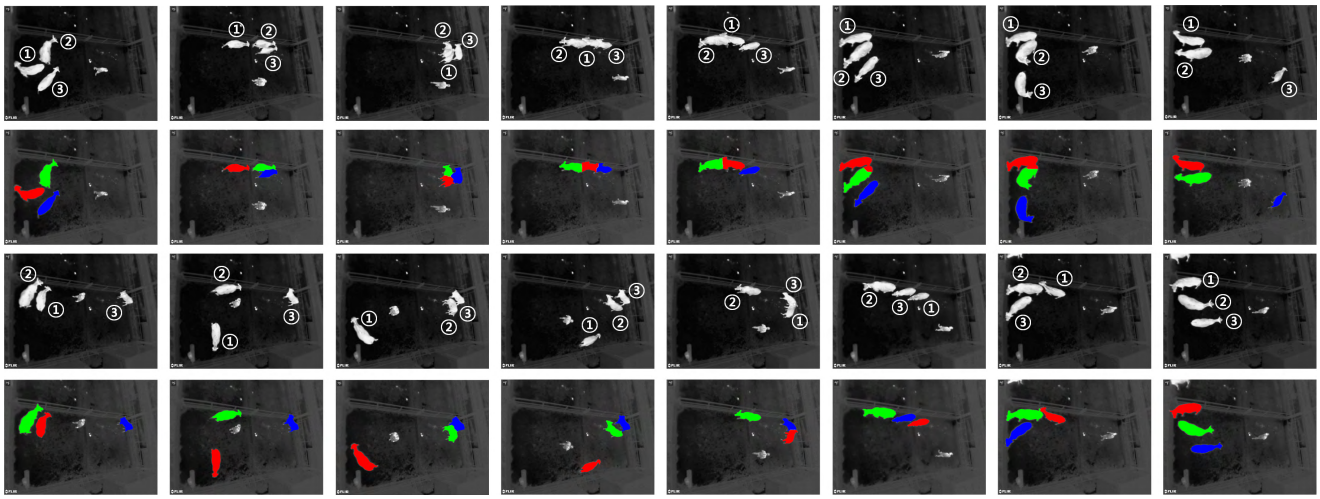


FIGURE 10. Complete results obtained by the proposed topographic surface segmentation using test sequence 1. Notice that ①, ②, and ③ are indicated in red, green, and blue, respectively. Best viewed in color.

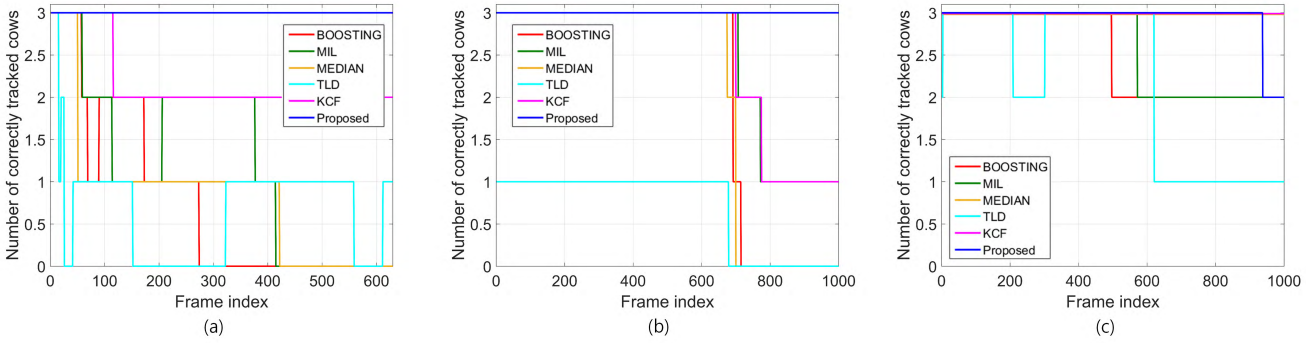


FIGURE 11. Comparison of the tracking performance based on the number of correctly tracked cows throughout the whole video sequences. (a) Results of Seq. 1. (b) Results of Seq. 2. (c) Results of Seq. 3.

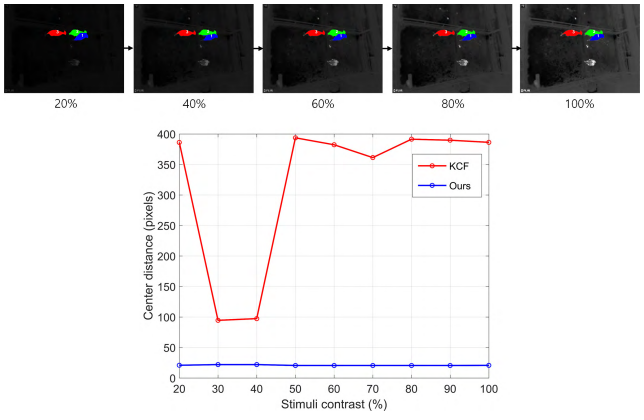


FIGURE 12. Performance according to the contrast of the input stimulus. Top : examples of tracking results of the proposed method with changing contrast. Bottom : Tracking accuracy in terms of the distance between estimated center positions and ground truth.

(C implementation). The processing times for all the methods employed for our experiments are shown in Table 4. Although certain approaches, e.g., MEDIAN [42] and KCF [27], are very fast, the proposed method provides tracking results that are reliable enough to be applied to real-time applications (> 30 fps). It should be noted that we can use multi-threaded image processing with a GPU to effectively increase the processing speed since the proposed algorithm can be applied to each object independently.

V. CONCLUSION

In this paper, a novel method for multiple-object tracking using a single thermal sensor has been proposed. The key idea of the proposed method is that each object captured from a top-down view can be interpreted as a topographic surface, which provides an accurate boundary even when multiple objects are overlapped. Due to the power of the thermal sensor, we can easily extract the foreground (i.e., animals) by a simple thresholding. Then, the topographic surface is generated based on a distance transform, which is subsequently fed into our segmentation scheme with makers (i.e., the center positions of multiple objects). The center position of each segmented object is consistently updated with a simple refinement scheme to be robust to the abrupt

motions of animals. Based on various experimental results, it is confirmed that the proposed method provides reliable tracking results for monitoring the behavior patterns of animals in the breeding space with a real-time processing speed.

REFERENCES

[1] 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.

[2] 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.

[3] K. Wang, Y. Liu, C. Gou, 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.

[4] C. Ma, Y. Wang, and G. Ying, "The pig breeding management system based on RFID and WSN," in *Proc. Int. Conf. Inf. Comput.*, Apr. 2011, pp. 30–33.

[5] L. Catarinucci *et al.*, "Smart RFID antenna system for indoor tracking and behavior analysis of small animals in colony cages," *IEEE Sensors J.*, vol. 14, no. 4, pp. 1198–1206, Apr. 2014.

[6] R. E. Floyd, "RFID in animal-tracking applications," *IEEE Potentials*, vol. 34, no. 5, pp. 32–33, Sep./Oct. 2015.

[7] P. Heas, C. Herzet, and E. Memin, "Bayesian inference of models and hyperparameters for robust optical-flow estimation," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1437–1451, Apr. 2012.

[8] M. A. Mohamed, H. A. Rashwan, B. Mertsching, M. A. Garcia, and D. Puig, "Illumination-robust optical flow using a local directional pattern," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 9, pp. 1499–1508, Sep. 2014.

[9] T. Crivelli, M. Fradet, P.-H. Conze, P. Robert, and P. Pérez, "Robust optical flow integration," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 484–498, Jan. 2015.

[10] G. Botella, A. Garcia, M. Rodriguez-Alvarez, E. Ros, U. Meyer-Baese, and M. C. Molina, "Robust bioinspired architecture for optical-flow computation," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 18, no. 4, pp. 616–629, Apr. 2010.

[11] G. Botella, J. A. M. H., M. Santos, and U. Meyer-Baese, "FPGA-based multimodal embedded sensor system integrating low- and mid-level vision," *Sensors*, vol. 11, no. 8, pp. 8164–8179, Aug. 2011.

[12] G. Botella, U. Meyer-Baese, A. García, and M. Rodríguez, "Quantization analysis and enhancement of a VLSI gradient-based motion estimation architecture," *Digit. Signal Process.*, vol. 22, no. 6, pp. 1174–1187, Dec. 2012.

[13] C. R. Maurer, R. Qi, and V. Raghavan, "A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 265–270, Feb. 2003.

[14] C. S. Regazzoni, "Distributed extended Kalman filtering network for estimation and tracking of multiple objects," *Electron. Lett.*, vol. 30, no. 15, pp. 1202–1203, Jul. 1994.

[15] S. A. Vigus, D. R. Bull, and C. N. Canagarajah, "Video object tracking using region split and merge and a Kalman filter tracking algorithm," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1. Oct. 2001, pp. 650–653.

[16] L. Marcenaro, M. Ferrari, L. Marchesotti, and C. S. Regazzoni, "Multiple object tracking under heavy occlusions by using Kalman filters based on shape matching," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3. Sep. 2002, pp. 341–344.

[17] J. MacCormick and A. Blake, "A probabilistic exclusion principle for tracking multiple objects," *Int. J. Comput. Vis.*, vol. 39, no. 1, pp. 57–71, 2000.

[18] K. Nummiaro, E. Koller-Meier, and L. Van Gool, "An adaptive color-based particle filter," *Image Vis. Comput.*, vol. 21, no. 1, pp. 99–110, Jan. 2003.

[19] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[20] J. G. Allen, R. Y. D. Xu, and J. S. Jin, "Object tracking using CamShift algorithm and multiple quantized feature spaces," in *Proc. Pan-Sydney Area Workshop Vis. Inf. Process.*, 2004, pp. 3–7.

[21] A. D. Jepson, D. J. Fleet, and T. F. El-Maraghi, "Robust online appearance models for visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1296–1311, Oct. 2003.

[22] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 125–141, May 2008.

[23] H. Grabner, M. Grabner, and H. Bischof, "Real-time tracking via on-line boosting," in *Proc. Brit. Mach. Vis. Conf.*, vol. 1. Sep. 2006, pp. 47–56.

[24] B. Babenko, M.-H. Yang, and S. Belongie, "Visual tracking with online multiple instance learning," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 983–990.

[25] S. Hare, A. Saffari, and P. H. S. Torr, "Struck: Structured output tracking with kernels," in *Proc. IEEE Conf. Comput. Vis.*, Nov. 2011, pp. 263–270.

[26] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1409–1422, Jul. 2012.

[27] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.

[28] A. Leykin, Y. Ran, and R. Hammoud, "Thermal-visible video fusion for moving target tracking and pedestrian classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[29] M. Talha and R. Stolkin, "Particle filter tracking of camouflaged targets by adaptive fusion of thermal and visible spectra camera data," *IEEE Sensors J.*, vol. 14, no. 1, pp. 159–166, Jan. 2014.

[30] C. Li, X. Sun, X. Wang, L. Zhang, and J. Tang, "Grayscale-thermal object tracking via multitask Laplacian sparse representation," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 4, pp. 673–681, Apr. 2017.

[31] 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.

[32] E. Gundogdu, A. Koc, B. Solmaz, R. I. Hammoud, and A. A. Alatan, "Evaluation of feature channels for correlation-filter-based visual object tracking in infrared spectrum," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun./Jul. 2016, pp. 290–298.

[33] A. Berg, J. Ahlberg, and M. Felsberg, "Channel coded distribution field tracking for thermal infrared imagery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun./Jul. 2016, pp. 1248–1256.

[34] J.-Y. Kwak, B. C. Ko, and J. Y. Nam, "Pedestrian tracking using online boosted random ferns learning in far-infrared imagery for safe driving at night," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 1, pp. 69–81, Jan. 2017.

[35] K. Z. Mao, P. Zhao, and P.-H. Tan, "Supervised learning-based cell image segmentation for P53 immunohistochemistry," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1153–1163, Jun. 2006.

[36] X. Yang, H. Li, and X. Zhou, "Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 11, pp. 2405–2414, Nov. 2006.

[37] S. Beucher and F. Meyer, "The morphological approach to segmentation: The watershed transformation," in *Mathematical Morphology in Image Processing*. New York, NY, USA: Marcel Dekker, 1993, pp. 433–481.

[38] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[39] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.

[40] S. Beucher and C. Lantuejoul, "Use of watersheds in contour detection," in *Proc. Int. Workshop Image Process.*, Sep. 1979, pp. 1–12.

[41] A. Mordvintsev and A. K, *OpenCV-Python Tutorials Documentation, Release 1*, Sep. 2017, pp. 143–149.

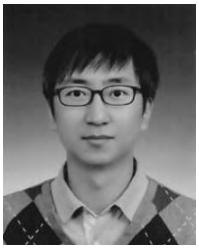
[42] Z. Kalal, K. Mikolajczyk, and J. Matas, "Forward-backward error: Automatic detection of tracking failures," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2756–2759.

[43] A. Berg, J. Ahlberg, and M. Felsberg, "A thermal object tracking benchmark," in *Proc. IEEE Int. Conf. Adv. Video Signal Surveill.*, Aug. 2015, pp. 1–6.

[44] H. B. Shitrit, J. Berclaz, F. Fleuret, and P. Fua, "Multi-commodity network flow for tracking multiple people," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1614–1627, Aug. 2014.

[45] S. Baysal and P. Duygulu, "Sentioscope: A soccer player tracking system using model field particles," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 7, pp. 1350–1362, Jul. 2016.

[46] Y. Li, C. Huang, and R. Nevatia, "Learning to associate: HybridBoosted multi-target tracker for crowded scene," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2953–2960.
[47] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *Int. J. Comput. Vis.*, vol. 12, no. 1, pp. 43–77, Feb. 1994.



WONJUN KIM (M'13) received the B.S. degree in electrical engineering from Sogang University, Seoul, South Korea, in 2006, the M.S. degree from the Department of Information and Communications, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2008, and the Ph.D. degree from the Department of Electrical Engineering, KAIST, in 2012. 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, 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, the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE SIGNAL PROCESSING LETTERS, and *Pattern Recognition*.

YONG BEOM CHO (M'86) received the B.Sc. degree from Kyongbuk University in 1981, the M.Sc. degree from the University of South Carolina in 1988, and the Ph.D. degree from Case Western Reserve University, Cleveland, OH, USA, in 1992. He was the Dean of International Office at Konkuk University, Seoul, South Korea, from 2006 to2009, where he is currently a Professor with the Department of Electronics Engineering. His research interests include embedded system design, SoC design, networking systems, the application of image processing to mobile environments, and digital communication system design for mobile and ad-hoc networks.

SANGRAK LEE received the B.Sc. and M.Sc. degrees from the Department of Feed Science, Konkuk University, Seoul, South Korea, in 1980 and 1985, respectively, and the Ph.D. degree from the Department of Animal Science, University of Tohoku, Sendai, Japan, in 1989. From 1989 to 1997, he was a Senior Researcher with the Animal Resources Research Center. Since 1997, he has been with the Department of Animal Science and Technology, Konkuk University, where he is currently a Professor. He was the Dean of the College of Animal Bioscience and Technology from 2012 to 2014 and the Dean of the Graduate School of Agriculture and Animal Science from 2013 to 2014.