LETTER

Moving Object Detection Based on Clausius Entropy

Jonghyun PARK[†a)], **Wanhyun CHO**[††], ***Nonmembers***, **Gueesang LEE**[††], ***Member***, ***and*** **Soonyoung PARK**[†], ***Nonmember***

SUMMARY This paper proposes a novel image segmentation method based on Clausius entropy and adaptive Gaussian mixture model for detecting moving objects in a complex environment. The results suggest that the proposed method performs better than existing methods in extracting the foreground in various video sequences composed of multiple objects, lighting reflections, and background clutter.

key words: *moving object detection, adaptive Gaussian distribution, Clausius entropy, foreground extraction*

1. Introduction

Detecting and tracking moving objects, representing the core technology of any surveillance system, have attracted considerable interest from computer vision researchers. However, because of dynamic changes in natural scenes, such as sudden illumination, weather changes, and repetitive motions causing clutter, reliable motion detection has been considered to be a difficult problem. Hence, the robustness of motion detection technology needs to be improved for applications in complex environments. Several techniques have been widely for detecting moving objects [1], [2]. The background subtraction method is a well-known technique for the motion segmentation of static scenes [3]. However, this technique is typically sensitive to dynamic changes when stationary objects uncover the background or when there are sudden illumination changes. In terms of typical statistical methods, Stauffer et al. [4] proposed an adaptive background mixture model for real-time tracking. In their model, every pixel is separately modeled by a mixture of Gaussians that are updated online by incoming image data. Another example of the statistical model is Rittscher et al.'s [5] probabilistic background model based on the hidden Markov model. This method can easily adapt to dynamic environments but has considerable difficulty extracting the complete shapes of certain types of moving objects. To overcome the shortcomings of two-frame differencing, improved methods have been proposed for three-frame differencing. For instance, Collins et al. [6] proposed a hybrid method that combines three-frame differencing with an adaptive background subtraction model for their VSAM project. Optical flow methods make use of flow vectors of moving objects over time to detect moving regions in an image. Barron et al. [7] compared the accuracy and reliability of different optical flow techniques for both real and synthetic image sequences. Several authors have proposed a new variational framework for detecting and tracking multiple moving objects in image sequences. Paragios et al. [8] considered Geodesic active contours or regions and level-set methods to address various tasks associated with optical flow estimation and to track moving objects in motion analysis. Bao et al. [9] presented a novel segmentation approach based on the spatial-temporal curve evolution framework for multiple moving objects. Lu et al. [10] proposed an improved motion detection method that integrates the temporal differencing method, the optical flow method, and the double-background filtering method. Tian et al. [11] proposed a real-time algorithm that can detect salient movements in complex environments by combing the temporal difference imaging and temporal filtered optical flow methods. E.J. Koh, et al. [13] proposed a motion segmentation method based on Clausius normalized field that can detect salient motion in complex scenes by Clausius entropy theory. The paper presents a new technique based on the Clausius entropy and background subtraction methods for detecting moving objects. The main goal of this algorithm is to more effectively separate the background from the foreground and detect moving objects accurately. First, we transform the initial region of moving objects into the Clausius entropy domain. Second, a technique for moving object detection is based on Clausius entropy. It is a background subtraction method that can model energy values in the entropy domain as finite Gaussian mixture.

Manuscript received March 8, 2010.
Manuscript revised October 12, 2010.

[†]The authors are with Mokpo National University, Chunggye, Muan, Chonnam, Korea.
[††]The authors are with Chonnam National University, Yongbong-dong, Buk-gu, Gwangju, Korea.
a) E-mail: jhpark@jnu.ac.kr

DOI: 10.1587/transinf.E94.D.388

2. Moving Object Detection Based on Clausius Entropy

2.1 Clausius Entropy

Entropy is a function of a quantity of heat in a system that is capable of doing work. Under maximum entropy, there is a minimum amount of energy available for work, whereas under minimum entropy, there is a maximum amount of energy available for work. Entropy S is not defined directly; it is defined by an equation reflecting changes in the entropy of the system as a result of changes in the heat of the system. A change in entropy (ΔS) is defined by

$$\Delta S = \frac{\Delta Q}{T} \tag{1}$$

where ΔQ is the amount of heat absorbed in an isothermal and reversible process in which the system goes from one state to another and T is the absolute temperature at which the process occurs [12]. If the temperature of the system is not constant, then this relationship is represented by the differential equation $dS = dQ/T$. To understand this equation, suppose that temperature T can be expressed as a function $T(Q)$ of heat Q. The total change in entropy is

$$\Delta S = \int_A \frac{1}{T(Q)} dQ, \tag{2}$$

where A is the set defining the range of heat values.

2.2 Computation of Clausius Entropy

The original Clausius entropy computation method was developed for motion detection and the equations in this section were derived by E.J. Koh et al. [13]. To compute Clausius entropy for pixel values in each frame of an image sequence, we need to define three items: the system or field F, energy or heat Q, and temperature T. First, we define the system or field F as an input video sequence I composed of gray-scales of a color image I_t. Each pixel in the frame image I_t of a video has a $W * W$ rectangular neighborhood or spatial window. To maintain Eq. (1), the system has to satisfy the following rules: (i) all responses are reversible; (ii) during a response, the system expends at the same temperature; and (iii) when the response is completed, the temperature changes before the next response. Second, we can define the absorbed energy from the input image I_t as follows:

$$\Delta Q_k^{(t)} = \sum_{l:all\ of\ pixels\ in\ window} w_k \left(X_{kl}^{(t)} - M_{kl}^{(t)}\right)^2, \tag{3}$$

where X_{kl} is the color value of the k^{th} channel at the l^{th} pixel in the window, $M_{kl}^{(t)}$ is the mean of all color values of the k^{th} channel for pixels, and w_k is the weight function for each channel. Moreover, to adapt the system to time, we have to update the mean value $M_{kl}^{(t)}$ for each frame image $I^{(t)}$. It is adjusted as follows:

$$M_{kl}^{(t)} = (1-\lambda)M_{kl}^{(t-1)} + \lambda \cdot X_{kl}^{(t)}, \tag{4}$$

where λ is the learning factor for adopting current means. Third, we define the absolute temperature of the system. At the microscopic level, temperature T can be defined as the average energy of each particle in the system. Hence, if we denote the proportional constant between heat Q and temperature T as κ, then the change in the temperature of the thermodynamic area can be defined as follows:

$$\Delta T = \kappa \frac{\Delta Q}{n}, \tag{5}$$

where n is the total number of particles belonging to some object. Here we can define the temperature in each frame as

$$T_k^{(t)} = (1-\rho)T_k^{(t-1)} + \kappa \frac{\Delta Q}{n}, \tag{6}$$

where ρ is a constant proportional to the amount of heat loss in every frame. Here temperature T satisfies the rule of the heat system, that is, the greater the difference in temperature between two objects, the greater the movement of energy is. Fourth, we can define the entropy variation for each channel as follows:

$$\Delta S_k^{(t)} = \frac{\Delta Q_k^{(t)}}{T_k^{(t)}} \tag{7}$$

Thus, we obtain the total entropy variations for each pixel (x, y) in the t^{th} frame by taking the sum of entropy variations in each channel:

$$\Delta S^{(t)}(x, y) = \sum_{k:channel} \Delta S_k^{(t)}(x, y) \tag{8}$$

Finally, the Clausius entropy method, based on the frame change in entropy, attempts to detect moving regions by making use of the sum of entropy variations for consecutive frames in a video sequence. It is well known that this method can be easily adapted to the static environment.

3. Background Substraction Model

Here we consider the changeable amount of the entropy of a particular pixel over time as a pixel process. At any given time t, what is known about a particular pixel (x, y) the history of its entropy variation:

$$\{X_1, \cdots, X_t\} = \{\Delta S^{(i)}(x, y) : \quad 1 \le i \le t\}, \tag{9}$$

where $\Delta S^{(i)}$ is the changeable amount of the entropy of the i^{th} frame image. The recent history of each entropy variation, $\{X_1, \cdots, X_t\}$, is modeled by a mixture of K Gaussian distributions [4], [14]. The probability of observing the current entropy variation at time t is

$$P(X_t) = \sum_{i=1}^{K} \xi_{i,t} \cdot \varphi\left(X_t; \mu_{i,t}, \sigma_{i,t}^2\right) \tag{10}$$

where K is the number of distributions, $\xi_{i,t}$ is the weight of the i^{th} Gaussian in the mixture $\varphi(\cdot)$, $\mu_{i,t}$ is the mean value, and $\sigma_{i,t}^2$ is the variance. The prior weight $\xi_{i,t}$ of K distributions at time t is adjusted as follows:

$$\xi_{i,t} = (1-\alpha)\xi_{i,t-1} + \alpha(M_{i,t}), \tag{11}$$

where α is the learning rate and $M_{i,t}$ is 1 for the distribution that match the new observation and 0 for other cases. In general, $\xi_{i,t}$ works as a low-pass filtered average of the posterior distribution that pixel values have matched model k given observations from 1 to t [4]. The parameters $\mu_{i,t}$ and $\sigma_{i,t}^2$ of the distribution that does not match the new observation maintain the same values. However, the parameters of the distribution that matches the new observation are updated as follows:

$$\begin{aligned} \mu_{i,t} &= (1-\rho)\mu_{i,t-1} + \rho \cdot X_t, \\ \sigma_{i,t}^2 &= (1-\rho)\sigma_{i,t-1}^2 + \rho(X_t - \mu_{i,t})^2 \end{aligned} \tag{12}$$

where $\rho = \alpha\varphi(X_t|\mu_{i,k}, \sigma_{i,k})$ works as the same type of the low-pass filter, except that only the data which matches the model is included in the estimation. Therefore, if an object is stationary just long enough to become part of the background and then it moves, the distribution describing the previous background still exists with the same μ and σ^2; however, it has lower ξ and is quickly reincorporated into the background [4]. We need to identify the Gaussian in the mixture that is most likely to be produced by background processes. The Gaussians are ordered by the value of ξ/σ. After the reestimation of the parameters of the mixture, it is sufficient to sort from the matched distribution to the most probable background distribution because only matched models' relative values would be changed. This ordering of models is effectively an ordered, open-ended list, where the most likely background distributions are found at the top and the less probable transient background distributions gravitate toward the bottom and are eventually replaced by new distributions. Then the first B distributions are chosen as the background model:

$$B = \arg\min_b \left(\sum\nolimits_{k=1}^{b} \xi_k > T\right), \tag{13}$$

where T is the measure of the minimum portion of the data that should be accounted for by the background. This process takes the best distributions until a certain portion, T, of the recent data can be accounted for. If T is low then the background model is likely to be unimodal. In this case, the most probable distribution will use to extract the background. If T is higher, then a multimodal distribution from a repetitive background motion can result in the inclusion of more than one color in the background model. This can in turn result in the transparency effect, which allows the background to accept two or more colors.

4. Experimental Results

We compared the performance of our method with that of the intensity-based method [4] for indoor/outdoor video sequences. Figure 1 shows the experimental results for the indoor video: (a) is the 445$^{\text{th}}$ frame of the input sequence; (b) is the entropy-dense image of our approach; (c) is the detection result based on the intensity component; and (d) is the detection result obtained using our method. Note that Fig. 1 (c) shows noise components because of lighting reflections of moving pedestrians on the well and the floor. However, as shown in Fig. 1 (d), the proposed method was able to detect moving objects without picking up any artifacts. Figure 2 shows the salient moving objects that were detected by the proposed method. The upper images (a)–(c) of Fig. 2 show the captured input sequence image, and the images (d)–(f) show the entropy-dense image generated by the Clausius entropy approach. The images (g)–(i) display noise components, and the images (h) and (i) do not perfectly segment the human region. However, as shown in the images (j)–(l), the proposed method was able to perfectly segment the images. Finally, to evaluate the performance of the proposed approach, we compared the segmentation precision ratios of the proposed and intensity-based methods. The precision ratio pixel is defined as

$$precision = \frac{N(S_m \cap S_a)}{N(S_m)} \times 100 \tag{14}$$

where $N(S_m)$ is the number of pixels in the ground truth and $N(S_m \cap S_a)$ denotes the number of identical pixels between the ground truth and the extracted foreground region. Figure 3 shows the performance of these methods. The av-

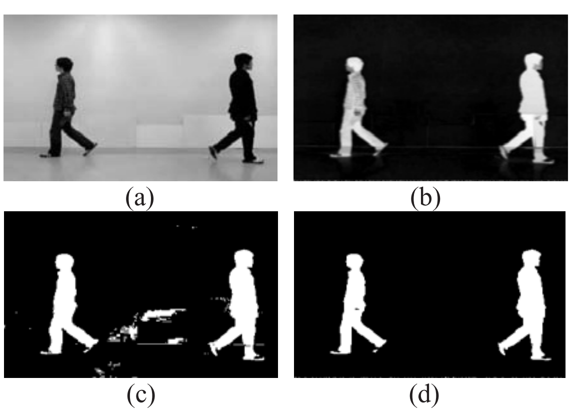

Fig. 1 Results for the indoor video: (a) the 445$^{\text{th}}$ input sequence; (b) the entropy-dense image by the proposed approach; (c) the detection result for the AGMM; (d) the detection result for the proposed approach.

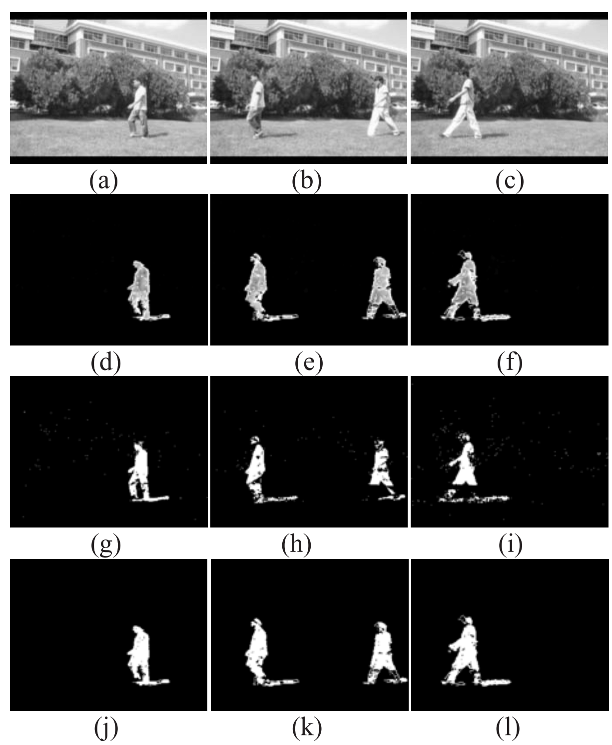

Fig. 2 Segmentation results: (a)–(c) input sequence images for the 240$^{\text{th}}$, 288$^{\text{th}}$, and 353$^{\text{rd}}$ frames; (d)–(f) dense maps generated by Clausius entropy for the 240$^{\text{th}}$, 288$^{\text{th}}$, and 353$^{\text{rd}}$ frames; (g)–(i) subtracted background images for the 240$^{\text{th}}$, 288$^{\text{th}}$, and 353$^{\text{rd}}$ frames; (j)–(l) the salient motion detected by the proposed method for the 240$^{\text{th}}$, 288$^{\text{th}}$, and 353$^{\text{rd}}$ frames.

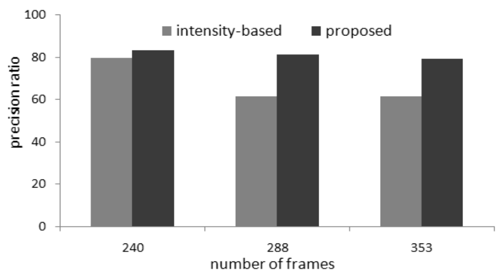

Fig. 3 Comparison of segmentation precision ratio to ground truth.

erage value of the precision rates of each frame for the proposed method was 81.36%. These results suggest that our method is more efficient in detecting moving objects than the intensity-based method.

5. Conclusions

The results suggest that the Clausius entropy method, together with the adaptive Gaussian mixture method, can substantially improve the detection of moving objects in video sequences. The proposed approach can transform the initial region of moving objects into the Clausius entropy domain, which can decompose images into the background under stable conditions and into the foreground under unstable conditions. The proposed method that can model entropy variations in the background as a mixture of Gaussians, is more reliable and robust than conventional methods in detecting moving objects both indoor and outdoor image sequences. Future research should develop more sophisticated methods that could detect moving objects in more complex scenes.

Acknowledgement

This work was partially supported by two National Research Foundations of Korea (NRF) grant funded by the Korean Government (MEST) (2010-0003360) and (KRF-2010-0013757). We would like to thank Eun Jin Koh for his kind help in providing us Clausius entropy computation method.

References

[1] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," Pattern Recognit., vol.36, pp.586–601, 2003.
[2] Y. Dedeoglu, Moving objects detection, tracking and classification for smart video surveillance, Master's Thesis, Depart. of Computer Eng., Bilkent University, Ankara, 2004.
[3] A.M. Mclvor, "Background subtraction techniques," Proc. Image and Vision Computing, Auckland, New Zealand, 2000.
[4] C. Stauffer and W.E.L. Grimson, "Learning patterns of activity using real-time tracking," IEEE Trans. Pattern Anal. Mach. Intell., vol.22, no.8, pp.747–757, 2000.
[5] J. Rittscher, J. Kato, S. Joga, and A. Blake, "A probabilitic background model for tracking," ECCV 2000, LNCS 1843, pp.336–350, 2000.
[6] R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, and O. Hasegawa, "A system for video surveillance and monitoring: VSAM final report," Technical Report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May 2000.
[7] J. Barron, D. Fleet, and S. Beauchermin, "Performance of optical flow techniques," Int. J. Comput. Vis., vol.12, pp.42–77, 1994.
[8] N. Paragios and R. Deriche, "Geodesic active regions and level sets methods for motion estimation and tracking," Computer Vision and Image Understanding, vol.97, pp.259–282, 2005.
[9] H. Bao and Z. Zhang, "Motion objects segmentation using a new level set based method," PCM 2004, LNCS 3331, pp.312–318, 2004.
[10] N. Lu, J. Wang, Q.H. Wu, and L. Yang, "An improved motion detection method for real-time surveillance," IAENG Int. J. Computer Science, vol.35, pp.1–16, 2008.
[11] Y.-L. Tian and A. Hampapur, "Robust salient motion detection with complex background for real-time video surveillance," IEEE Workshop on Motion and Video Computing (WACV/MOTION'05), vol.2, pp.30–35, 2005.
[12] P. Pierre, A to Z of Thermodynamics, Oxford University Press, 1998.
[13] E.J. Koh, J.H. Won, and C.S. Bae, "Clausius normalized field for motion segmentation," Proc. IPIU Conference, 2009.
[14] C. Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real time tracking," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.246–252, 1999.