Image Segmentation in Thermal Images

Yung-Yao Chen*, Wei-Sheng Chen, and Hui-Sheng Ni
Graduate Institute of Automation Technology
National Taipei University of Technology
Taipei, Taiwan
yungyaochen@mail.ntut.edu.tw*

***Abstract*—This paper presents a novel method for image segmenting in thermal images. A thermal imaging sensor is a device that forms an image by detecting the infrared radiation from objects in a scene and comparing their temperature differences. In this study, we characterized an object's heat signature (its mean temperature and variance) by segmenting its contours in thermal images and analyzing the information from each foreground segment. However, because of the limitations of thermal sensors (e.g., a high signal-to-noise ratio and low contrast), manually contouring thermal images is difficult and time consuming. We propose a semiautomatic segmentation method that takes advantage of information from another optical camera. Such a bi-sensor system improves the precision of contour sketching for objects in thermal images. The experimental results demonstrated the robustness of the proposed method.**

Keywords—image segmentation; image rectification; K-means clustering; thermal imaging; interactive approach

I. INTRODUCTION

Thermal imaging sensing technology can normally detect radiation in the long-infrared range of the electromagnetic spectrum [1]. This is because of the law of black-body radiation; every object emits infrared radiation, as long as its temperature is above absolute zero. A thermal imaging sensor provides superior visibility in comparison to typical digital cameras in dark environments. Recently, thermal imaging sensors have found widespread use in applications such as surveillance [2], electronic packaging [3], and intelligent traffic systems [4]. In this study, we characterize the heat (i.e., temperature) signature of different objects. As the reason, thermal imaging sensor is implemented to obtain the complete temperature information of each point on the object; however, successfully segmenting an object in a thermal image remains difficult.

The goal of image segmentation is to separate desired objects from the background, and is an essential task in computer vision and object recognition [5-7]. Generally, color and texture information are used for segmentation. However, in thermal images the only information available is the variation in temperature, which makes segmentation difficult. Moreover, fully automatic segmentation in a natural image is extremely complex. Semiautomatic segmentation methods that require a few simple user inputs would be superior alternatives.

In this paper, we propose a novel method for image segmenting in thermal images. We integrated a semiautomatic segmentation called maximal similarity-based region merging (MSRM) [8] into our system. Because over-segmentation (i.e., segmenting the input image into too many segments) is the main challenge for MSRM, we proposed a segmentation method that can be considered for the preprocessing of MSRM. The auxiliary information of another digital camera is used to handle the over-segmentation problem, to successfully segment the object from the background in a thermal image. The remainder of this paper is organized as follows. In Sec. II, we briefly describe the proposed method. In Sec. III, we provide the experimental results. Finally, we draw our conclusions in Sec. IV.

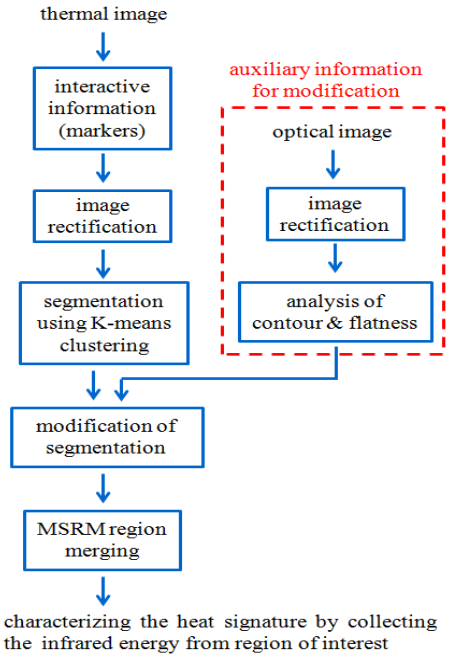


Fig. 1. Overall framework of the proposed thermal image segmentation algorithm.

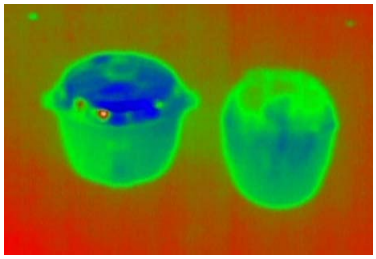
II. PROPOSED METHOD

Before we describe the proposed method in details, we first introduce the thermal imaging sensor used in this study. Fig. 2(a) shows the thermographic camera cores (model: Tau 2) [9], and Figs. 2(b) and 2(c) show two example images from this camera: color mode and grayscale mode, respectively. All objects emit infrared radiation as a function of their heat energy (temperature). Generally, the higher temperature an object is, the more radiation it emits. Rainbow (color) display mode is the most popular display mode in thermal imaging

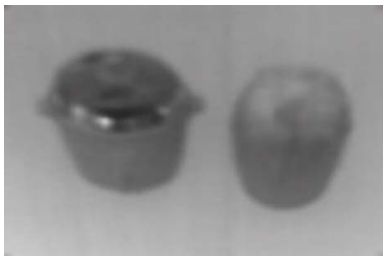
Supported in part by the National Science Council under Grants NSC MOST 104-2221-E-027-032.

application, however, it has been post-processing (adding color) to help users easily identify objects at different temperatures. In natural, thermal images are normally grayscale: cold objects are black, hot objects are white, and the gray intensity indicates variations between the two. In this study, grayscale thermal images are used.

(a)



(b)



(c)

Fig. 2. Thermal imaging sensor used in this study. (a) Long-wave infrared camera (Tau 2); (b) Color display mode; and (c) Grayscale display mode.

A. Image Rectification

Because of the limitations of components and sensing abilities, compared to optical sensor technologies, the common drawbacks of thermal imaging are: 1) a higher signal-to-noise ratio; 2) a lower spatial and brightness resolution; and 3) lower contrast and blurring. Such drawbacks make automatic (or semiautomatic, as described in Section 3) image segmenting outside the region of interest in the thermal image difficult. Therefore, we established a bi-sensor system by adding information from an optical image of the same scene.

The goal of image rectification is to map a point on an original image plane to the rectified image plane. By performing the rectification process twice, we can integrate relevant information from both thermal and optical images, and project them onto a common rectified image plane. The concept of image rectification is briefly described as follows. First, the homogeneous coordinate vectors $\vec{Q}$ and $\vec{q}$ corresponding to a three-dimensional (3-D) real world coordinate and a 2-D image coordinate, respectively, are defined as

$$\vec{Q} = [X \quad Y \quad Z \quad 1]^T, \tag{1}$$

and

$$\vec{q} = [x \quad y \quad 1]^T. \tag{2}$$

The relationship between a point in the real world coordinate and its corresponding location in the image coordinate that captured by a camera, can be constructed by

$$\vec{q} = sMW\vec{Q}, \tag{3}$$

where s is the scaling matrix, M is the intrinsic camera parameter matrix, and W is the extrinsic camera parameter matrix. This assumes that for each point in the real world, its corresponding thermal and optical image locations can be mapped onto the same image plane by the projective transformation in (3). The assumption holds when the thermal imaging sensor and the digital camera are placed close together and the camera-to-subject distance is within a certain range. Equation (3) can be simplified as

$$\vec{q} = H\vec{Q}, \tag{4}$$

where H is normally referred to as a homography matrix in projective geometry that transforms an point onto different projective coordinates. Derived from some matrix calculation in (4), the mapping (H_t) from a point in the thermal image onto a rectified image plane and the mapping (H_o) from a point in the optical image onto the same rectified image plane can be expressed by

$$\vec{q}_{rect} = H_t\vec{q}_t, \tag{5}$$

and

$$\vec{q}_{rect} = H_o\vec{q}_o, \tag{6}$$

where $\vec{q}_t$ and $\vec{q}_o$ represent the homogeneous coordinate vectors in the thermal image coordinate and optical image coordinate, respectively; and $\vec{q}_{rect}$ represent the homogeneous coordinate vector in the rectified image. The homography matrix H is a 3×3 matrix defined by

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix}, \tag{7}$$

where the eight unknown entries of H indicate the way of projection. Applying (7) to (5), the mapping can be expressed as

$$\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_{rect} \\ y_{rect} \\ 1 \end{bmatrix}, \tag{8}$$

where $(x_{rect}/z_{rect}, y_{rect}/z_{rect})$ and (x_t, y_t) indicate the actual coordinates of the point in the rectified image and point in the thermal image, respectively. Given a point in the thermal image, we can easily identify its counterpart in the rectified image. Further expanding the above (8), the mapping can be expressed by

$$\begin{cases} x_{rect} = h_{11}x_t + h_{12}y_t + h_{13} \\ y_{rect} = h_{21}x_t + h_{22}y_t + h_{23} \\ z_{rect} = h_{31}x_t + h_{32}y_t + z_t \end{cases}, \tag{9}$$

which implies that the corresponding 2-D coordinates (x', y') of the point in the rectified image are

$$x' = \frac{x_{rect}}{z_{rect}} = \frac{h_{11}x_t + h_{12}y_t + h_{13}}{h_{31}x_t + h_{32}y_t + z_t}, \tag{10}$$

and

$$y' = \frac{y_{rect}}{z_{rect}} = \frac{h_{21}x_t + h_{22}y_t + h_{23}}{h_{31}x_t + h_{32}y_t + z_t}, \tag{11}$$

The above formula represents two equations which need to be solved to find H. Because there are eight unknowns in H, we need four such point pairs $\{(x_i', y_i'), (x_{t,i}, y_{t,i}), i = 1,2,3,4\}$ to find the value of H, i.e.,

$$\begin{bmatrix} x_{t,1} & y_{t,1} & 1 & 0 & 0 & 0 & -x_{t,1}x_1' & -y_{t,1}x_1' \\ 0 & 0 & 0 & x_{t,1} & y_{t,1} & 1 & -x_{t,1}y_1' & -y_{t,1}y_1' \\ x_{t,2} & y_{t,2} & 1 & 0 & 0 & 0 & -x_{t,2}x_2' & -y_{t,2}x_2' \\ 0 & 0 & 0 & x_{t,2} & y_{t,2} & 1 & -x_{t,2}y_2' & -y_{t,2}y_2' \\ x_{t,3} & y_{t,3} & 1 & 0 & 0 & 0 & -x_{t,3}x_3' & -y_{t,3}x_3' \\ 0 & 0 & 0 & x_{t,3} & y_{t,3} & 1 & -x_{t,3}y_3' & -y_{t,3}y_3' \\ x_{t,4} & y_{t,4} & 1 & 0 & 0 & 0 & -x_{t,4}x_4' & -y_{t,4}x_4' \\ 0 & 0 & 0 & x_{t,4} & y_{t,4} & 1 & -x_{t,4}y_4' & -y_{t,4}y_4' \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} = \begin{bmatrix} x_1' \\ y_1' \\ x_2' \\ y_2' \\ x_3' \\ y_3' \\ x_4' \\ y_4' \end{bmatrix}. \tag{12}$$

Equation (12) can be written in the form of $A\vec{x} = \vec{b}$, where A is an 8×8 coefficient matrix, $\vec{x}$ consists of the eight unknowns of H, and $\vec{b}$ contains the constant terms. The least mean square solution (LMS) of $\vec{x}$ can be expressed as

$$\vec{x} = \left(A^T A\right)^{-1} A^T \vec{b}, \tag{13}$$

Replacing the values in $\vec{x}$, the homography matrix is obtained. Independently repeating (8) to (13) twice, a point in both the thermal and optical images can be projected onto the same rectified image. Fig. 3 shows a rectified optical image that corresponds to Fig. 2(c).

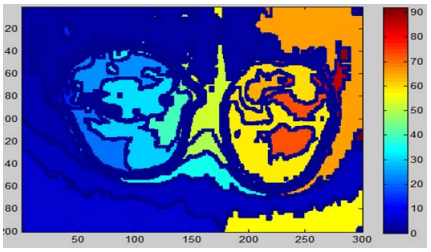
Fig. 3. Example of a rectified optical image.

B. Initial Segmentation by K-Means Clustering

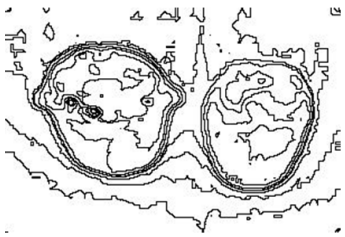
In this phase, initial image segmentation is performed in the grayscale thermal image (Fig. 2(c)) as follows.

- Construct a feature space based on the pixel value (i.e., temperature information).
- Classify all samples in the feature space into K clusters by using the K-means algorithm. The value of K is set within the range of three to six times the desired number of groups. For example, in the case of two foreground objects and one background, the desired number is three.
- Label each pixel in the thermal image using the K-means result.

Fig. 4(a) shows the thermal image labeled by a cluster index based on temperature information. Because the difference of the spatial coordinate is considered, the number of segments is actually much greater than the value of K. Hence, the segmentation result is further modified by introducing auxiliary information from the corresponding rectified optical image.



(a)



(b)

Fig. 4. Segmentation of a thermal image using the K-means clustering algorithm ($K = 10$). (a) Image labeled by the cluster index; and (b) Results of image segmentation.

C. Analysis of Contours and Flatness

In this phase, two features of the rectified optical image are extracted: the contours and flatness. Image segmentation using only the information from a thermal image alone might cause problems because of the common limitations of thermal imaging sensing mentioned in Section 2A. In particular, low contrast and blurring make directly contouring a thermal image difficult. Therefore, the first data we want to obtain from the optical image is contour information.

The contour analysis consists of two parts: 1) edge enhancements, and 2) edge detection. In this study, they were performed during the spatial domain processing. For edge enhancements, local discontinuities at the boundaries of different objects were enhanced as follows.

- Convolve a Gaussian smoothing operator with the original image.
- Subtract a fraction of the smoothed image from the original.
- Boost the amplitude of all frequency components for edge enhancement.

After enhancing the edges of the original image, contour information is obtained using Sobel operators to extract the edge pixels form the enhanced image. The operators are defined as

$$G_{normal} = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \tag{14}$$

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \tag{15}$$

and

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}, \tag{16}$$

where (14) is the Gaussian smoothing operator, and (15) and (16) are the Sobel operators in the horizontal and vertical directions, respectively. Fig. 5(a) shows an example of contour analysis, in which the black pixels are labeled as the contour pixels. From Fig. 5(a), the contours of the two objects are clear; however, the noise from the mottled background also produces some fake contours that must be removed.

The flatness analysis consists of two parts and is based on spatial domain processing. First, bilateral filtering is performed to reduce the noise of the image while preserving the edges. Both the Euclidean distance of the pixels and the difference of (color) pixel values are considered. The bilateral filter can be expressed as

$$I_{\mathbf{p}}^{filtered} = \frac{1}{W_{\mathbf{p}}}\sum_{\mathbf{q}\in S} G_{\sigma_s}\left(\|\mathbf{p}-\mathbf{q}\|\right) G_{\sigma_r}\left(\left\|I_{\mathbf{p}} - I_{\mathbf{q}}\right\|\right) I_{\mathbf{q}}, \tag{17}$$

where $\mathbf{p}$ and $\mathbf{q}$ are certain 2D pixel positions, S defines the neighborhood of $\mathbf{p}$, $I_{\mathbf{p}}^{filtered}$ is the filtered pixel value at $\mathbf{p}$, $I_{\mathbf{q}}$ is the original pixel value at $\mathbf{q}$, G_{σ_s} is the spatial Gaussian weighting (for smoothing differences in coordinates), and G_{σ_r} is the range Gaussian weighting (for smoothing differences in intensities). The normalization term $W_{\mathbf{p}}$ is used to ensure that the filter preserves the image's average gray value:

$$W_{\mathbf{p}} = \sum_{\mathbf{q}\in S} G_{\sigma_s}\left(\|\mathbf{p}-\mathbf{q}\|\right) G_{\sigma_r}\left(\left\|I_{\mathbf{p}} - I_{\mathbf{q}}\right\|\right), \tag{18}$$

Fig. 5(b) shows an example of a bilateral filtered image, in which, compared to Fig. 3, the sharp edges are preserved and the tiny color differences (i.e., noise) between the pixels of local areas in the background are alleviated. The next step is to determine the flat regions of the bilateral filtered image, which is achieved by analyzing a cost function ϕ that represents the change in intensity for the shift (u,v):

$$\phi(u,v) = \sum_{x,y\in S} w(x,y)\left[I(x+u,y+v) - I(x,y)\right]^2, \tag{19}$$

where ϕ is a window function with a Gaussian distribution, S defines the window size, and $I(x,y)$ and $I(x+u,y+v)$ represent the pixel values at (x,y) and $(x+u,y+v)$, respectively. In other words, ϕ indicates the weighted sum of the squared differences (SSD) between the two patches of

$$patch_1 = \sum_{x,y\in S} I(x+u,y+v), \tag{20}$$

and

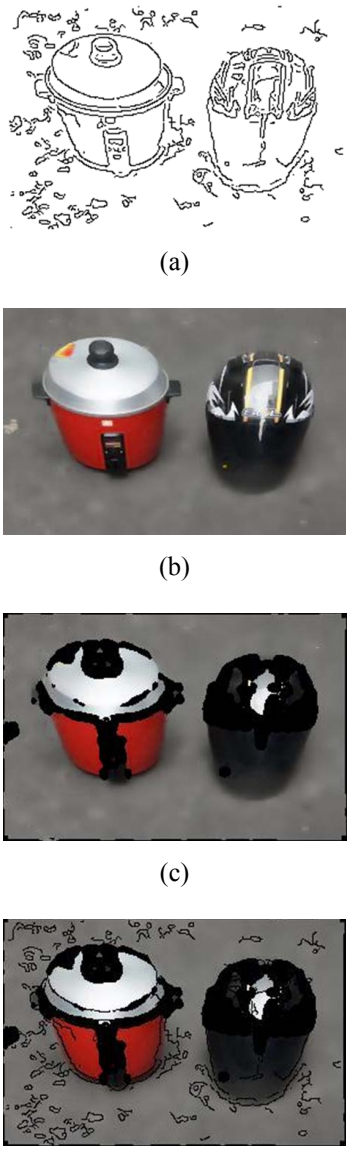
$$patch_2 = \sum_{x,y\in S} I(x,y), \tag{21}$$

If ϕ is small for the case of all directions in shifts (u,v), the pixel at (x,y) is considered a flat pixel. To simplify the calculation, the SSD can be approximated by a first-order Taylor expansion, i.e.,

$$I(x+u,y+v) \approx I(u,v) + I_x(u,v)x + I_y(u,v)y, \tag{22}$$

where I_x and I_y are the partial derivatives of I in the horizontal and vertical directions, respectively. It was derived in [10] that by applying (22) to (19), ϕ can be written in matrix form, and the value of ϕ can be quickly identified using an eigen-value analysis. Fig. 5(c) shows an example flatness analysis result, in which black pixels are overlaid to

show the flat regions. Fig. 5(d) shows the integrated information output of this phase (from the rectified optical image), in which we have labeled both contour pixels and flat pixels on the bilateral filtered image.



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

Fig. 5. Analysis of the rectified optical image. (a) Contours; (b) Bilateral filtered image; (c) Flat pixels (overlaid as black); and (d) Integrated information.

D. Segment Merging Using Optical Image Information

In this phase, the initial segmentation of the thermal image is modified using information from the rectified optical image (see Section 2C); the goal of modification, basically, is to merge some of the segments shown in Fig. 4(b) on the basis of optical information. Segment merging in this phase involves two steps. First, non-contour pixels are merged with each other using the minimum color difference criteria. Each of the non-contour pixels is visited in a raster order (from left to right and top to bottom). At each processing position, its cluster index is renewed according to one of the indices of the eight neighboring pixel positions. To determine the new cluster index, the color differences between the current pixel and its eight neighboring pixels are calculated. If both the current and one of the neighboring pixels are labeled as flat pixels, the cluster index is renewed to be the same as the cluster index of the neighboring pixel (i.e., merging them together). If one of the neighboring pixels is labeled as a contour pixel, the color difference from the current pixel is not calculated (i.e., they are not eligible for merging). Otherwise, we simply compare the eight color differences and assign a new cluster index to the neighboring pixel with the minimum color difference, as shown in Fig. 6(a).

After merging the segments that contain non-contour pixels, non-visited pixels (i.e., contour pixels) are then compulsorily merged using a predefined merge rule. These pixels are similarly visited in a raster order. Fig. 6(b) shows an example of using a predefined merge rule to renew the cluster index of the current pixel. The only constraint to the merge rule is that at least one of its eight neighboring pixels must have been processed in the previous merging. In the beginning, it might be possible that a certain pixel and all of its eight neighbors are contour pixels (i.e., no cluster indices have been renewed). In such a case, the pixel position is skipped but is revisited in the next iteration.

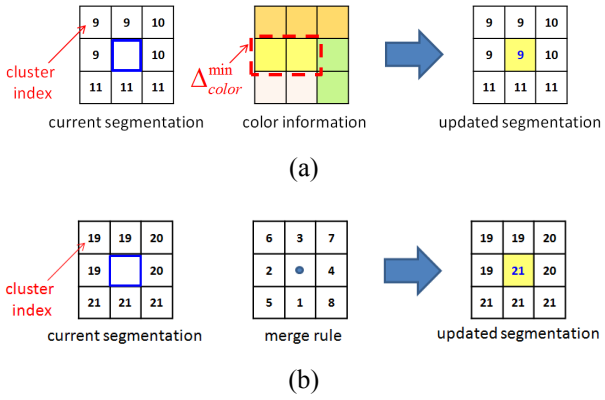
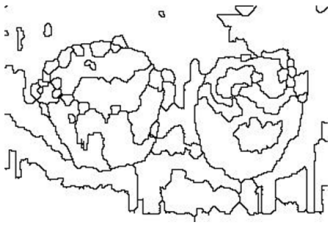


Fig. 6. Segment merging using optical image information. (a) Merging of non-contour pixels; and (b) Merging of contour pixels.

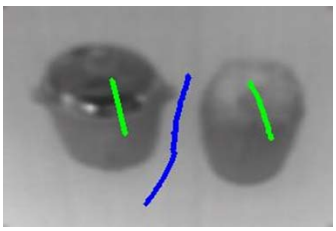
III. EXPERIMENTAL RESULTS

In this study, a novel method for image segmentation in thermal images is proposed as a preprocessing step in the interactive MSRM method proposed in [8]. In [8], the image is first segmented by a mean-shift approach, and then merged to the desired number of segments by using MSRM. However, because of the abovementioned limitations of thermal sensing, the mean-shift segmentation approach, as well as other segmentation approaches designed for optical images, is unsuitable for thermal images. The main cause for error is the excess tiny segments inputting into the MSRM; however, the low contrast of thermal images makes comparing the similarity between different segments difficult. Therefore, some features extracted from the rectified optical image were used to merge some of these tiny segments. Fig. 7(a) shows the results of the modified segmentation; compared to Fig. 4(b), the number of segments is effectively reduced. Fig. 7(b)

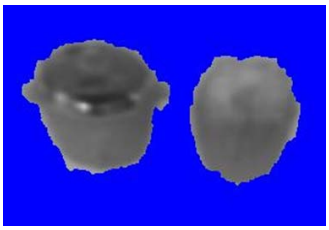
shows the marks input by a user, in which the green line is an object marker and the blue line indicates the background. Fig. 7(c) shows the final output of the MSRM, in which the blue pixels represent the background. The experimental result demonstrates that combining the proposed method with the MSRM method can effectively contour the objects in a thermal image.



(a)



(b)



(c)

Fig. 7. Experimental results. (a) Modified segmentation; (b) User inputs; and (c) Final segmentation result after MSRM, in which the blue pixels represent the background pixels.

IV. CONCLUSION

Thermal imaging has many advantages compared to optical imaging, such as improving the visibility of objects under low luminance and the detection of object temperature. However, image segmentation in thermal images is difficult because of the limitations of thermal imaging abilities. Conventional image segmentation approaches designed for optical images are not suitable for thermal images. This paper presents a novel method that can be applied for thermal image segmentation. The experimental results show that the proposed method can be successfully integrated as a pre-processing step of the MSRM method.

REFERENCES

[1] N. Nandhakumar and J.K. Aggarwal, "Integrated analysis of thermal and visual images for scene interpretation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 469-481, 1988.

[2] W. Wang, J. Zhang, and C. Shen, "Improved human detection and classification in thermal images," *IEEE Int. Conf. Image Processing*, pp. 2313-2316, 2010.

[3] W. Chen, H. Cheng, and H. Shen, "An effective methodology for thermal characterization of electronic packaging," *IEEE Trans. Components and Packaging Technologies*, vol. 26, pp. 222-232, 2003.

[4] A. Sangnoree, and K. Chamnongthai, "An application of thermography for its to evaluate the over limited velocity vehicles in nighttime-traffic," *IEEE Int. Conf. ITS Telecommunications*, pp. 150-153, 2008.

[5] *M.M.S.J. Preetha, L.P. Suresh, and M.J. Bosco, "Image segmentation using seeded region growing", IEEE Int. Conf. Computing, Electronics and Electrical Technologies (ICCEET)*, pp. 576–583, 2012.

[6] P.K. Jain and S. Susan, "An adaptive single seed based region growing algorithm for color image segmentation", *IEEE India Conference (INDICON)*, pp. 1–6, 2013.

[7] R. Gothwal, S. Gupta, D. Gupta, and A.K. Dahiya, "Color image segmentation algorithm based on RGB channels", *IEEE Int. Conf. Reliability, Infocom Technologies and Optimization (ICRITO)*, pp. 1–5, 2014.

[8] J. Ning, L. Zhanga, D. Zhanga, and C. Wub, "Interactive image segmentation by maximal similarity based region merging", *Pattern Recognition*, vol. 43, issue 2, pp. 445–456, 2010.

[9] Available online. http://www.flir.tw/home/.

[1] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Proc. 4th Alvey Vision Conference*, pp. 147-151, 1988.