第33卷第1期
2014年2月

红外与毫米波学报
J. Infrared Millim. Waves

Vol. 33, No. 1
February, 2014

文章编号：1001－9014(2014)01－0106－11

DOI：10.3724/SP.J.1010.2014.00106

Level set based segmentation of moving humans in thermal infrared sequences

GUO Yong-Cai[1], TAN Yong[1,2*], GAO Chao[1]

(1. Key Laboratory of Optoelectronic Technology and Systems of the Education Ministry, Chongqing University, Chongqing 400030, China;
2. Physics and Electronic Engineering Department, Yangtze Normal University, Chongqing 408003, China)

Abstract: The level set based active contour model (LSAC) has been proved advantageous for image segmentation. Based on LSAC techniques, a novel algorithm was proposed to overcome the difficulties of image segmentation in infrared human detection systems. It consists of a motion-based LSAC module, a threshold-based LSAC module and a fusion module. The motion-based LSAC, which bridges level set and background-subtraction techniques, conducts foreground segmentation and background estimation simultaneously based on converged level set functions. It works for detecting the moving regions in a sequence. Moreover, its output is regarded as the input of the threshold-based LSAC, which combines level set and thresholding techniques. This threshold-based LSAC module has the ability to extract the image regions having intensities within the range specified by dual thresholds and works for detecting all possible regions that may contain human candidates. Finally, the third module fuses the LSAC outputs and results in faithful segmentation result owing to the morphological open reconstruction. Furthermore, the fast numeric scheme proposed for evolving the LSAC modules and the optimized algorithmic flow improves efficiency. Experimental results demonstrate that the algorithm enjoys better performance in accuracy, efficiency and robustness to camera movement and temporal changes in the scene in comparison with the rival algorithms.

Key words: image processing; level set based active contour; thermal infrared human image; open reconstruction

PACS: 07.05.Pj

基于水平集的热红外运动人体目标分割算法

郭永彩[1], 谭 勇[1,2*], 高 潮[1]

(1. 重庆大学 光电工程学院光电技术及系统教育部重点实验室,重庆 400030;
2. 长江师范学院 物理学与电子工程学院,重庆 408003)

摘要: 水平集活动轮廓模型是一种优秀的图像分割方法. 针对红外人体检测系统中的图像分割难题,提出了一种基于水平集活动轮廓模型的新算法. 该算法包含水平集运动检测模块、水平集亮度检测模块和融合模块. 水平集运动检测模块融合了水平集和背景相减技术,通过演化水平集函数同时实现前景分割和背景估计,它用于检测序列中的运动区域,并将其演化结果输入到下一检测模块. 水平集亮度检测模块融合了水平集和阈值分割技术. 在给出双阈值时,可分割出亮度在双阈值所限定范围内图像区域,它用于检测序列图像序列中可能包含人体目标的全部区域. 利用形态学开重建技术,融合模块在融合前两个模块检测结果后输出算法最终分割结果. 此外,采用快速数值算法演化水平集检测模块以及优化设置整个算法流程,改善算法运行效率. 实验结果表明,相对其他典型算法,该算法具有较高分割精度和运行效率,且对时序亮度变化和镜头运动鲁棒性更好.

关 键 词: 图像处理;水平集活动轮廓模型;热红外人体图像;开重建

中图分类号: TN911.73 **文献标识码**: A

Received date: 2012-10-31, **revised date**: 2013-12-12 **收稿日期**: 2012-10-31, **修回日期**: 2013-12-12

Foundation items: Supported by Ph. D. Programs Foundation of Ministry of Education of China (20090191110026); Chinese Fundamental Research Funds for the Central Universities (CDJXS11120025)

Biography: Guo Yong-Cai (1963-), female, Chongqing, professor, Ph. D. Research area involves signal processing and pattern recognition. E-mail: ycguo@cqu.edu.cn.

* **Corresponding author**: E-mail: cquty@126.com.

Introduction

Thermal Infrared imaging is independent of illumination and works efficiently in dark and poor lighting conditions. Owing to this advantage, human detection in infrared imagery has been widely used in the systems for military night vision, traffic security, etc.. Image segmentation, which separates objects or events of interest from their surroundings, works as a fundamental step. However, perfect segmentation is hardly available due to the following reasons. Firstly, the poor image quality, such as image blur, poor resolution and clarity, low foreground/background contrast and heavy noises, makes it difficult to detect the objects in infrared images, although such objects may be brighter than the background and seldom affected by such factors as light changing, shadows and clothes. Secondly, many disturbing objects can also be captured by infrared cameras due to their hot temperature. Thirdly, great variations of poses, sizes, body shapes and appearance make it hard to extract faithful human silhouettes and complete interiors.

Many methods have been proposed for infrared human image segmentation. Among them, thresholding [1] may be the most popular one due to its simplicity and efficiency, but it often causes serious fragmentation. The method called projection[2] has similar advantages but it just adapts to simple human patterns. Motion-dependent techniques[3] are also proposed for the task. Unfortunately, they can not completely avoid the deficiencies such as fragmentation, and sometimes they work slowly.

Active contour models are generally impractical for infrared human image segmentation due to their inefficiency. However, the trials of applying them to infrared images can be attractive. Firstly, they can achieve sub-pixel accuracy of object boundaries. Secondly, they provide a flexible energy minimization framework in which multiple image cues can be naturally integrated. Thirdly, enclosed, smooth contours can be directly presented. Early models are usually difficult to handle topological changes of the contour until the level set method[4] was proposed. It represents implicitly the active contour as the zero level set of a higher dimensional function called level set function (LSF), and deforms this function instead of directly evolving the contour to approach object boundary. Thereafter, many level set based active contour models (LSACs) have been proposed.

To overcome the difficulties in the segmentation of infrared human sequences, a novel level-set based algorithm was presented. In comparison with rival methods, the proposed algorithm is advantageous in segmentation accuracy, efficiency and the robustness to camera motions and temporal changes in the scene.

The rest of the paper is organized as follows: in section 1, level set basics are introduced. In section 2, the proposed algorithm is presented in detail. Experimental results and analysis are given in section 3. Finally, some conclusions are drawn in section 4.

1 Level set basics

Let $C(\boldsymbol{p},t)$, defined as $\{x(\boldsymbol{p},t),y(\boldsymbol{p},t)\}$, denote a time-dependant curve starting from the initial position $C_0(\boldsymbol{p})$. The motion of $C(\boldsymbol{p},t)$ is governed by the equation written as

$$\begin{cases}\dfrac{\partial C(\boldsymbol{p},t)}{\partial t} = F(\mathscr{K})\boldsymbol{N} \\ C(\boldsymbol{p},t=0) = C_0(\boldsymbol{p})\end{cases}, \quad (1)$$

where $F(\mathscr{K})$ is a function concerning about the mean curvature $\mathscr{K}$. It defines the moving speed of such dynamic curve in the direction of its Euclidean normal inward vector $\boldsymbol{N}$. The level set method is an efficient numerical technique for curve evolution. In this method, a scalar Lipschitz function $\phi(\boldsymbol{p},t)$, i. e., the LSF, embeds an n-dimensional surface $\boldsymbol{S}$ in an R^{n+1} space surface. The points on surface $\boldsymbol{S}$ are mapped by $\phi(\boldsymbol{p},t)$ such that $S=\{\boldsymbol{p}\mid\phi(\boldsymbol{p},t)=c\}$, where c is an arbitrary scalar. In other words, $\boldsymbol{S}$ is the c level set of the function ϕ. Actually, the zero level set is often treated as the curve $C(\boldsymbol{p},t)$. In accordance with (1), the first order of partial differential equation (PDE) is represented as

$$\frac{\partial\phi}{\partial t} = F\mid\nabla\phi\mid, \quad (2)$$

where $\mid\nabla\phi\mid$ represents some appropriate finite different operators for the spatial derivative, and ∂t the temporal step. Since different topologies of the zero level

set do not imply the different topologies of the level set function ϕ, this LSF is supposed to be topology free, which facilitates the tracking of the curve within the evolution.

2 Algorithmic details

The proposed algorithm consists of three modules, i. e., the motion-based LSAC module for detecting moving human candidates in a sequence, the threshold-based LSAC module for detecting all the regions that may contain human candidates in the sequence, and the fusion module refining outputs of previous modules to produce algorithmic results that are faithful to ground-truths. Note the LSFs used in the LSAC modules are denoted by the symbol ϕ with some subscripts.

2.1 The Motion-Based LSAC

Let f_t denote the frame captured by an infrared camera at time t. Considering the sequence slice that consists of the frames $f_{t-n}, n = 1,2,\ldots,N$, possible temporal changes or camera motions may be slight in this slice. This situation can be especially true when such slice lasts pretty short time. Therefore, an object that moves across the slice can be given by background subtraction with an assumed background image. In this section, the background image is denoted by B. Let Ω be the image domain of the frame f_i, $\forall i \in [t-N,t]$, and C_{mi} be the active contour that separates f_i into interior region $inside(C_{mi})$ (i. e., the object) and exterior region $outside(C_{mi})$ (i. e., the background). Let $\phi_{mi}: \Omega \rightarrow R$ be the LSF that implicitly represents the partition of Ω as follows:

$$\begin{cases} C_{mi} = \{(x,y) \mid \phi_{mi}(x,y) = 0\} \\ inside(C_{mi}) = \{(x,y) \mid \phi_{mi}(x,y) > 0\} \\ outside(C_{mi}) = \{(x,y) \mid \phi_{mi}(x,y) < 0\} \end{cases}$$

the following energy functional can be minimized to detect the object that moves across such slice:

$$E_{mi} = \iint_{\Omega} F(\phi_{mi}(x,y))(a - (f_i(x,y) - B(x,y))^2)dxdy + \mu_1 \int_{C_{mi}} ds, \forall i \in [t-N,t] \quad , \quad (3)$$

where $F(\phi_{mi})$ is a function defined as

$$F(\phi_{mi}) = 0.5 + \sin\frac{\phi_{mi}}{6\beta}, \phi_{mi} \in [-\beta\pi, \beta\pi], \forall i \in [t-N,t] \quad . \quad (4)$$

It can be seen that F lies in $[1/2, 1]$ if $0 \leqslant \phi_{mi} \leqslant \beta\pi$ and $[0, 1/2]$ if $-\beta\pi \leqslant \phi_{mi} < 0$. β is a positive parameter that controls the upper and lower bounds of ϕ_{mi} variation, and a is another parameter working for the differencing between f_i and the background B.

The second term on the right hand side of Eq. (3) is a arc length related regularization term, which assures the smoothness of C_{mi} and avoids the occurrences of small isolated regions. The parameter μ_1 controls the weight of this term in the whole functional.

Using the variational principal, the gradient descent flow equation corresponding to Eq. (3) is as follows:

$$\frac{\partial\phi_{mi}}{\partial\tau} = \underbrace{\frac{1}{6\beta}\cos\frac{\phi_{mi}}{6\beta}((f_i - B)^2 - a)}_{Data\text{-}dependant\ term} + \underbrace{\mu_1 \cdot \delta(\phi_{mi}) div\left(\frac{\nabla\phi_{mi}}{|\nabla\phi_{mi}|}\right)}_{Curvature\text{-}dependant\ term}, \forall i \in [t-N,t] \quad , \quad (5)$$

where τ is artificial time, $\delta(.)$ is the *Dirac* function defined by $\delta(z) = \partial H(z)/\partial z$, in which $H(z)$ takes '1' if $z \geqslant 0$ and '0' if $z < 0$.

Examining Eq. (5), only the first term on its right hand side depends on image data. Given a point $\boldsymbol{p}$, this data-dependant term is positive if $(f_i - B)^2(\boldsymbol{p}) > a$ and negative if $(f_i - B)^2(\boldsymbol{p}) < a$, since the cosine function is strictly non-negative. Correspondingly, the LSF ϕ_{mi} would increase when this term is positive, or decrease when negative. As a result, the contour C_{mi} expands to get $\boldsymbol{p}$ included in interior region, or shrinks to get $\boldsymbol{p}$ excluded from interior region. In this way, all the points where $(f_i - B)^2(\boldsymbol{p}) > a$ can be satisfactorily grouped into object regions, and the left group into background regions. The parameter a can be viewed as a detection parameter, as it decides the points where the absolute differences between f_i and B are significant. In this paper, a is empirically determined by $a = 4r^2$, where r is the mean of the absolute-difference matrix elements that are bigger than the median of such matrix. The second term concerns about the mean curvature $\mathcal{K}$, which is equal to the divergence of ϕ_{mi}. This curvature-dependant term guides the contour C_{mi} to move at speed $\mathcal{K}$ on the image domain Ω. Since spurious regions have bigger mean curvature, the motion results in the elimination of the spurious regions and the smoothness of the contour. Guided by the data-

dependant term and the curvature-dependant term, the contour C_{mi} evolves to form smooth object boundary and makes the image segmentation when ϕ_{mi} converges.

Minimizing the total energy of the slice $E = \sum_{i=t-N}^{t} E_{mi}$ with respect to B, and then setting the gradient to be zero, it produces

$$B = \frac{\sum_{i=t-N}^{t}\iint_{\Omega} F(\phi_{mi}(x,y))f_i(x,y)\,dxdy}{\sum_{i=t-N}^{t}\iint_{\Omega} F(\phi_{mi}(x,y))\,dxdy} \quad . \quad (6)$$

As can be seen from Eq. (6), the background image B is the weighting average of $(N+1)$ frames, and $F(\phi_{mi})$ works as a weighting function. Since the foreground corresponds to larger values of $F(\phi_{mi})$ than the background does, Eq. (6) makes the update of background image constrained by last segmentation result. Discretizing the spatial partial derivatives $\partial\phi_{mi}/\partial x$, $\partial\phi_{mi}/\partial y$, and the temporal partial derivative $\partial\phi_{mi}/\partial\tau$, the solution to Eq. (5) ϕ_{mi}^{*}can be iteratively searched by

$$\phi_{mi}^{k+1} = \phi_{mi}^{k} + \Delta t.\, L(\phi_{mi}^{k}) \quad , \quad (7)$$

where Δt is the temporal step, and $L(\phi_{mi}^{k})$ is the discretized terms on the right hand side of Eq. (5) in kth iteration. Eq. (7) converges so slowly that the following numeric scheme utilizing a binary level set function (BLSF) [5] is adopted. The detailed steps of this scheme are as follows:

(1) Initialize ϕ_{mi}^{0} as a binary function, which takes 1 and -1 in the regions inside and outside of the contour, respectively. Then, set the stopping criterion $T = |\phi_{mi}^{k+1} - \phi_{mi}^{k}| < \lambda$, where λ is a very small constant working as a threshold;

(2) Calculate ϕ_{mi}^{k+1} by $\phi_{mi}^{k+1} = \phi_{mi}^{k} + \Delta t.\, V$, where V is the discretized data-dependant term in Eq. (5);

(3) Get the sign of ϕ_{mi}^{k+1} and smooth it by a Gaussian filter, i.e., $\phi_{mi}^{k+1} = sign(\phi_{mi}^{k+1}) * G_{\sigma}$, where G_{σ} is the Gaussian filter with standard deviation σ. The symbol "$*$" denotes a convolution operator;

(4) Repeat step (2) ~ (3) until it meets the stopping criterion T. Finally, output $\phi_{mi}^{*} = sign(\phi_{mi}^{k+1})$.

Note the size of G_{σ} used in above procedure can be truncated to $K \times K$, where K is the smallest positive integer that is bigger than 4σ. A truncated Gaussian filter with the size 5×5 can be seen in Fig. 1.

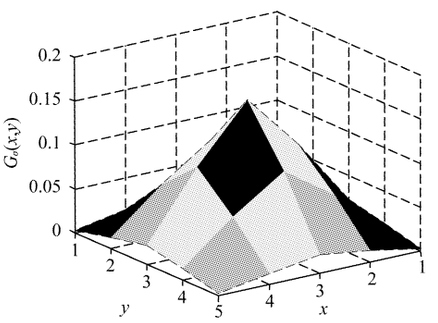


Fig. 1　A truncated 2-D Gaussian filter ($K=5, \sigma=1$)
图 1　截断的二维高斯滤波器

2.2 The Threshold-Based LSAC

Since humans usually appear brighter than the background in an infrared image, they can be simply and effectively detected by thresholding. However, incandescent objects, such as light bulbs and vehicles, which appear brighter than humans, can often be observed in surroundings. Therefore, it is appropriate to set dual threshold values for a thresholding algorithm. Let L and U, $0 \leqslant L < U \leqslant 255$ denote such dual threshold values. Based on the work in Ref. 1, the following criteria are proposed to decide L and U:

$$\begin{cases} L = \frac{5}{4}\varepsilon(\gamma + \sigma') \\ U = \min(3 \times \frac{5}{4}\gamma, 255) \end{cases} \quad , \quad (8)$$

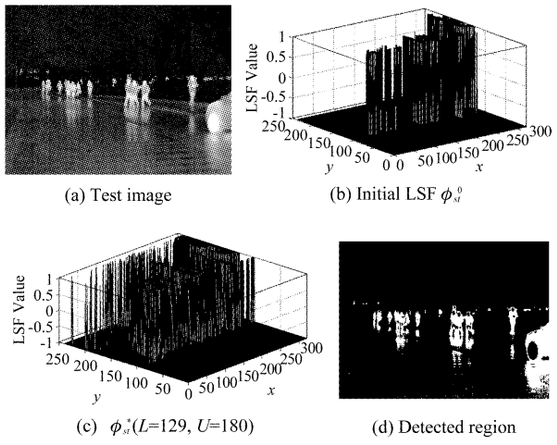
where γ and σ' are the average and standard deviation of the frame f_t. respectively. The parameter ε varies in $(0, 1]$. Image contrast and noise level should be considered for the choice of ε. Afterwards, a quadratic function is defined as follows:

$$g(z) = \left(\frac{L-U}{2}\right)^2 - \left(z - \frac{L+U}{2}\right)^2 \quad . \quad (9)$$

As can be seen, this function is positive if z lies between (L, U) and negative if it lies beyond the range. Based on $g(z)$, the regions having intensities within $[L, U]$ can be got by minimizing this energy functional

$$E_{st} = -\iint_{\Omega} H(\phi_{st}(x,y))g(f_t(x,y))\,dxdy + \iint_{\Omega} H(-\phi_{st}(x,y))g(f_t(x,y))\,dxdy + \mu_2\int_{C_{st}} ds \quad , (10)$$

where $\phi_{st}$is the LSF, C_{st} is the active contour (or the zero level set of ϕ_{st}).



(a) Test image (b) Initial LSF ϕ_{st}^0

(c) $\phi_{st}^*(L=129, U=180)$ (d) Detected region

Fig. 2 Exemplar images for the threshold-based LSAC
图2 基于阈值的LSAC模型运行效果演示图

The minimum of Eq. (7) would be met when all pixels with the intensities within (L, U) are grouped. In other words, by minimizing Eq. (7), all regions that may contain humans in the frame f_t can be achieved. Using the variational principal, the gradient descent flow equation corresponding to Eq. (10) is as follows:

$$\frac{\partial \phi_{st}}{\partial \tau} = \delta(\phi_{st})\left(g(f_t) + \mu_2 div\left(\frac{\nabla \phi_{st}}{|\nabla \phi_{st}|}\right)\right) \quad (11)$$

The solution ϕ_{st}^* can also be sought by the numeric scheme given in section 2.1. An illustration for this LSAC is given in Fig. 2. As can be seen from the segmentation result Fig. 2 (d), the background and the incandescent wheel get excluded simultaneously.

2.3 The Fusion Module

Based on the solution ϕ_{mt}^* to (5) and the solution ϕ_{st}^* to (11), the foreground regions in f_t can be given by

$$\begin{cases} R_{mt} = \{(x,y) \mid \phi_{mt}^*(x,y) > 0\} \\ R_{st} = \{(x,y) \mid \phi_{st}^*(x,y) > 0\} \end{cases}$$

Owing to the factors such as image noise, human inhomogeneity, occlusion and disturbing objects in surroundings, human regions in R_{mt} and R_{st} may be fragmentary and spurious regions may occur. For example, Fig. 3 (b) is the segmentation result of Fig. 3 (a) by the motion-based LSAC, and Fig. 3 (c) the result by the threshold-based LSAC. One sees that human fragments and spurious regions (corresponding to image noise and the bright lamp) occur. To eliminate spurious regions, open reconstruction (OREC)[6] that entirely keeps connected component shapes is more applicable than regular morphological filters that may result in jagged boundaries. OREC starts from a marker (or seed) and spreads in flood-fill fashion to recover subregions of a mask in which the marker lies. Fig. 4 illustrates the mechanism of this operation.

Spurious regions in the LSAC outputs do not locally correspond due to the random of image noise and the disturbing objects detected by a single LSAC module. This situation is different for human regions which overlap completely or partially. So, one can always get the markers by finding the overlapped regions in human regions but not in spurious regions. By starting from these markers and then recovering the connected components in which the markers lie, spurious regions in each of the results achieved by LSAC modules get removed while human regions get recovered with no loss of boundary smoothness. Furthermore, The OREC outputs are fused by an OR operation. In this way, human fragments merge to be meaningful regions. This process can be expressed as

$$R_t = R_{st}^{rec} \cup R_{mt}^{rec} = ((R_{mt} \cap R_{st}) \Delta R_{st}) \cup ((R_{mt} \cap R_{st}) \Delta R_{mt}) \quad (12)$$

where "Δ" denotes the OREC operator. Note the *AND* result between R_{mt} and R_{st} works as the marker, and R_{st}, R_{mt} work as the mask, respectively. Fig. 3 (d) shows the fusion result between Fig. 3 (b) and Fig. 3 (c). It can be seen that most spurious regions disappear, and in the meanwhile, human interior fragments get reduced dramatically. Fig. 3 (e) shows human contours before and after the fusion. Clearly, the red contour derived from the fusion module fits the human silhouette much better than the blue or green contour derived from a single LSAC module. What's more, no loss of contour smoothness can be found by examining Fig. 3 (f), since the red one always overlaps on certain pieces of the blue or the green one.

2.4 The Complete Work Flow

Figure 5 shows the working flow of the proposed algorithm. Except first $(N+1)$ initial LSFs, which should be manually set for motion segmentation of the first $(N+1)$ frames, the left initial LSFs which correspond to left frames are automatically given. The background image B can be simply initialized as the mean, or the median of the first several frames of the se-

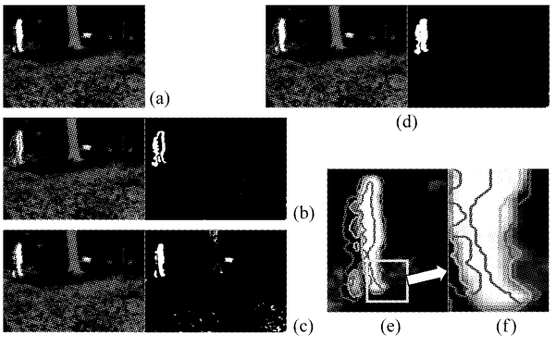


Fig. 3　Exemplar images to illustrate the fusion module. (a) Original image. (b) Result by the motion-based LSAC (Left: final contour. Right: interior region). (c) Result by the threshold-based LSAC. (d) Fusion result. (e) Human contours before (blue and green lines) and after the fusion (red line). (f) Zoomed view of the rectangle region in (e)
图3　融合模块运行效果演示图. (a) 原图像. (b) 运动LSAC模块分割结果(左: 最终活动轮廓曲线. 右: 目标内部区域). (c) 基于阈值的LSAC模块分割结果. (d) 融合结果. (e) 融合前人体轮廓(蓝色与绿色曲线)与融合后的人体轮廓(红色曲线). (f) 图(e)中矩形区域的放大视图

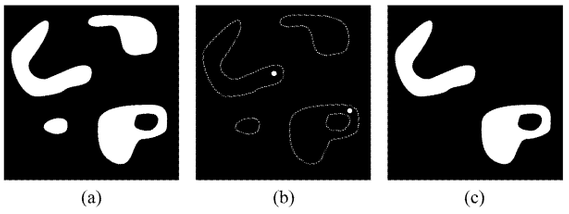


Fig. 4　Exemplar images to illustrate open reconstruction. (a) The mask. (b) The maker (bright regions). (c) Result by open reconstruction
图4　开重建的运行效果演示图. (a) 掩膜. (b) 标记(高亮区域). (c) 开重建结果

quences, or even the first frame.

For each frame, the output of the motion-based LSAC is used as the initial LSF for the threshold-based LSAC to make the threshold-based LSAC converge rapidly since the initial contour is close to object boundary.

In the step of LSF updating, only one new LSF is actually computed for the new coming frame while the others are sequentially obtained from previous computations. This situation is similar in the step of background updating. Exemplar images to demonstrate the working flow can be seen in Fig. 6.

3　Experimental Results and Analysis

The proposed algorithm (PRO) was implemented by Matlab7.1 on a computer with Intel Core Duo 1.66 GHz CPU, 2G RAM, and Windows XP operating system. Three thermal infrared clips, which provide relatively easy, moderate, and stiff conditions for human segmentation respectively, were used for testing the PRO. In clip 1#, one subject enters the field of view (FOV) from the left. The subject is then hidden behind a tree for about 19 seconds, and continues walking right. In clip 2#, the subject enters the FOV from the right and continues walking left to a door at the end of a hallway where it exits the building. Clip 3# was acquired by our FLIR A40 thermal infrared camera on campus. The subjects enter the FOV on the top left corner and walk along the road. Since the camera pans in a clockwise direction horizontally, the subjects exit the FOV on the bottom left corner. More details about the clips are given in Table 1. Note the heading frames of each clip are shown in top row, and the overlaid arrows on the frames images indicate the motion direction of the subjects. Unless otherwise specified, the parameters used for the PRO are as follows: $N=2, \beta=6$, $\Delta t=5$, $\lambda=1\times10^{-6}$, $\varepsilon=1$ for clip 1# ~ 2# and $\varepsilon=0.85$ for clip 3#, owing to their differences in contrast and noise conditions. The size of Gaussian filter G_σ with $\sigma=1$ is 5×5. The LSF provided for the motion-dependant LSAC module is initialized as ϕ_0, and Fig. 7 shows its 3-D and top views. Moreover, the background image is initialized as the first frame of each clip. For comparison, the mixtures of Gaussians (MOG) [7] is chosen as a representative of statistical methods, Lee's method [8] as a representative of background-subtraction based level set methods, and Li's method [9] as a representative of frame-differencing based level set methods. Three Gaussian kernels and adaptive thresholding are used in the MOG. The background images used in Lee's method are initialized by the median method. All parameters with the rival methods are tunable for achieving as better results as possible.

The results in clip 1# ~ 3# are shown in Figs. 8 ~ 10, respectively. In Fig. 8, the top row shows four representative frames sequentially labeled as S_1, S_2, S_3 and S_4. Moving human regions are highlighted in the frames by white rectangles, and their motion directions are indicated by arrows. Four middle rows show the results by the referred methods above. The red

contours overlapping on the frames denote the final contours derived by the methods. The regions inside red contours are the detected objects. Besides, the bottom row shows the manually labeled ground-truths. Fig. 9 and Fig. 10 are configured in the same way. Moreover, the threshold values used in the threshold-based LSAC are listed in Table 2.

Examining Fig. 8, the MOG based result is rather fragmentary due to an occlusion that the subject undergoes soon after it enters the screen from the left. Lee's and Li's methods perform better than the MOG but their results are still incomplete. The PRO based result is least affected by occlusion, because its fusion module eliminates as many fragments as possible.

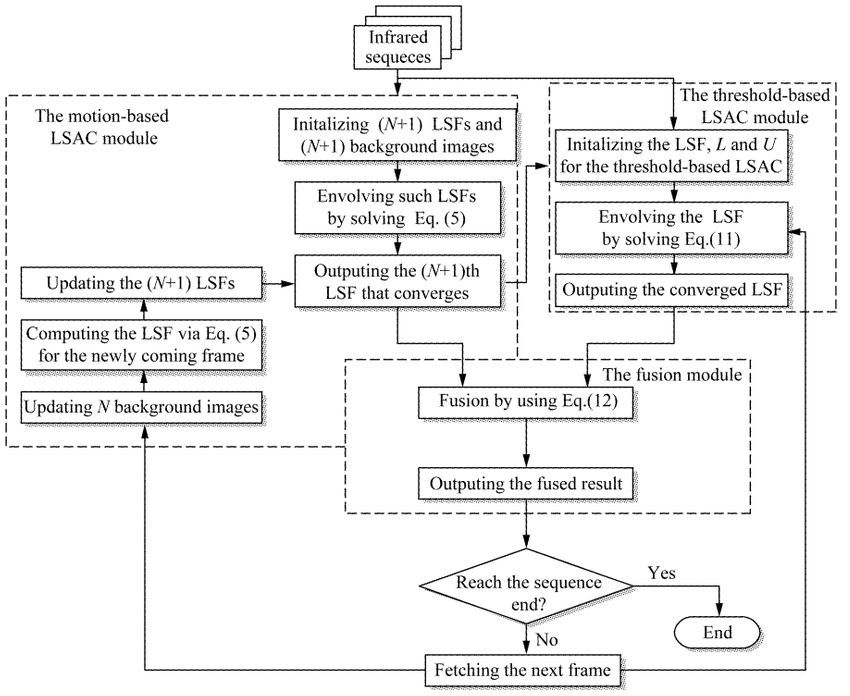


Fig. 5 The working flow of the proposed algorithm
图5 本文算法运行流程

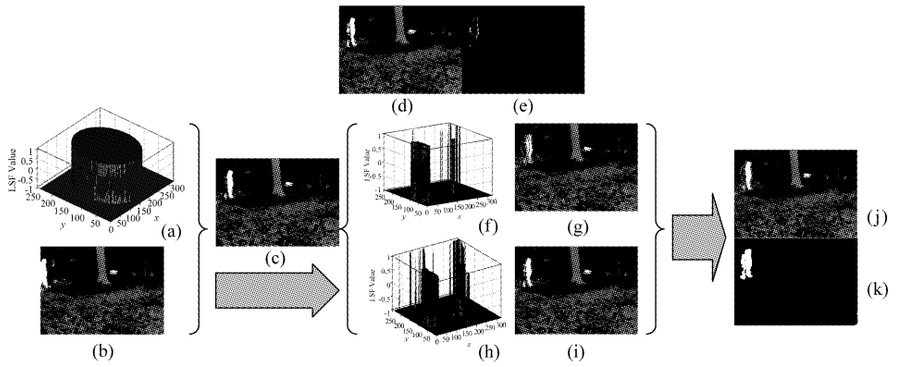
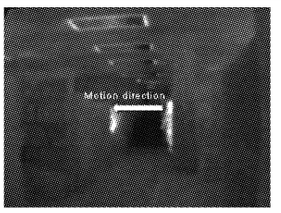


Fig. 6 Exemplar images to demonstrate the working flow. (a) The initial LSF manually set for the motion-based LSAC, (b) Initial background image, (c) The frame to be segmented, (d) The estimated background for (c), (e) Absolute difference between (c) and (d), (f) The converged LSF of the motion-based LSAC, (g) Object contour derived from (f), (h) The converged LSF of the threshold-based LSAC, (i) Object contour derived from (h), (j) Object contour given by the complete algorithm, (k) Object interior
图6 本文算法工作流程演示图.(a) 运动LSAC模块的LSF初始值;(b) 初始背景参考图像;(c) 待分割的帧图像;(d) 图(c)的背景估计图像;(e)图(c)和图(d)的绝对差值图像;(f) 运动LSAC模块输出的解;(g) 由(f)得到的目标轮廓;(h) 基于阈值的LSAC模块输出的解;(i) 由(h)得到的目标轮廓;(j) 完整算法所得目标轮廓;(k) 目标内部区域

Table 1　Details of test clips
表 1　测试序列详细信息

		
Clip 1#: Outdoor clip including a subject with relatively clear boundary and homogeneous interior. Selected from "OTCBVS Infrared Benchmark". Good contrast. Acquired by fixed Raytheon L-3 Thermal-Eye 2000AS camera. Frame size = 320 × 240 pixels. 8-bit grayscale JPEG format. 100 frames.	Clip 2#: Indoor clip including a subject with blurred boundary and homogeneous interior. Selected from "OTCBVS Infrared Benchmark". Acquired by fixed Raytheon L-3 Thermal-Eye 2000AS camera. Frame size = 320 × 240 pixels. 8-bit grayscale JPEG format. 80 frames.	Clip 3#: Outdoor clip with some shadowing, intensive noise, and the subjects having strong inhomogeneities. Shot by FLIR A40 thermal infrared camera. FOV = 24°. Distance to target D ≤ 100 m. Panning rate ≈ 4.5°/sec. 15 frames/sec. 8-bit grayscale BMP format. Frame size = 320 × 240 pixels, 68 frames.

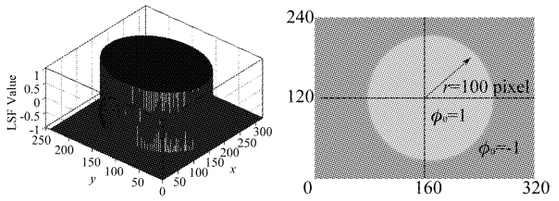


Fig. 7　The initial LSF ϕ_0. Left: 3D view. Right: top view
图 7　LSF 初始值 ϕ_0. 左: 立体视图. 右: 顶视图

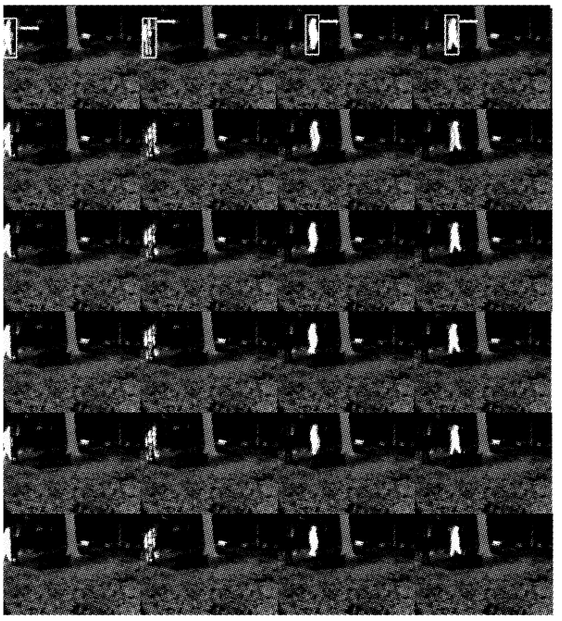


Fig. 8　Segmentation results in clip 1#. First row: Sample frames labeled as $S_1 \sim S_4$ sequentially. Second row: result by the MOG. Third row: result by Lee's method. Fourth row: result by Li's method. Fifth row: result by the PRO. Sixth row: the ground-truth
图 8　PRO 与各对比方法在 clip 1#上的分割结果. 第一行: 依次编号为 $S_1 \sim S_4$ 的序列帧. 第二行: MOG 方法结果. 第三行: Lee 方法结果. 第四行: Li 方法结果. 第五行: PRO 方法结果. 第六行: 真实人体区域

Later, all methods except the MOG give faithful foreground regions as soon as the occlusion disappears, owing to relatively good contrast of the whole clip and intensity homogeneity of the subject.

Examining Fig. 9, the PRO misclassifies the fewest pixels belonging to human region into the background. So, the subject region can be more faithful to ground-truths than those given by its rivals. Also, it misclassifies the fewest background pixels into the foreground. For example, no human regions should be found in S_4 (positioning at the tail in top row) since the person has exited. However, spurious regions can still be seen in the rival based results.

Clip 3# presents great challenge to the algorithms. As can be observed from Fig. 10, the MOG and Lee's methods present meaningless segmentation results, as they are intrinsically dependent on relatively static background image. The PRO and Li's methods provide much better results. In comparison with Li's method, the PRO misclassifies fewer background pixels, and makes the result "clearer".

Jaccard similarity and the mean absolute distance (*MAD*)[10] are chosen as the metrics to quantitatively evaluate above results. The Jaccard similarity is defined as $S_J = |A_S \cap A_M| / |A_S \cup A_M|$, where A_S, A_M are extracted foreground region and the ground-truth, respectively. One can see a perfect overlap between A_S and A_M if $S_J = 1$ and no overlap if $S_J = 0$.

The *MAD* measures the dissimilarity between the extracted boundary and the ground-truth. A smaller *MAD* value means that the extracted boundary is closer

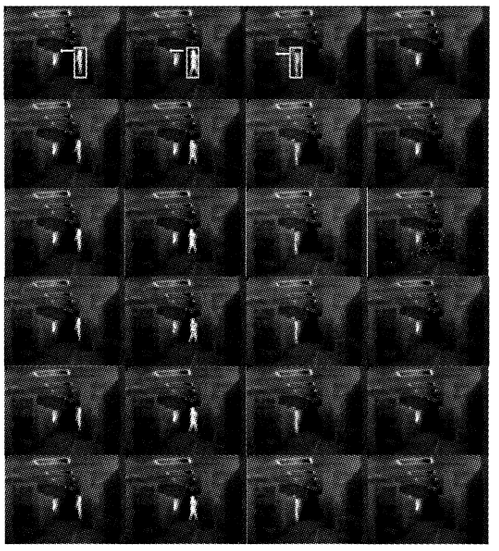


Fig. 9 Segmentation results in clip 2#. First row: Sample frames labeled as $S_1 \sim S_4$ sequentially. Second row: the result of the MOG. Third row: the result by Lee's method. Fourth row: result by Li's method. Fifth row: the result by the PRO. Sixth row: the ground-truth

图9 PRO与各对比方法在clip 2#上的分割结果. 第一行: 依次编号为 $S_1 \sim S_4$ 的序列帧. 第二行: MOG方法结果. 第三行: Lee方法结果. 第四行: Li方法结果. 第五行: PRO方法结果. 第六行: 真实人体区域

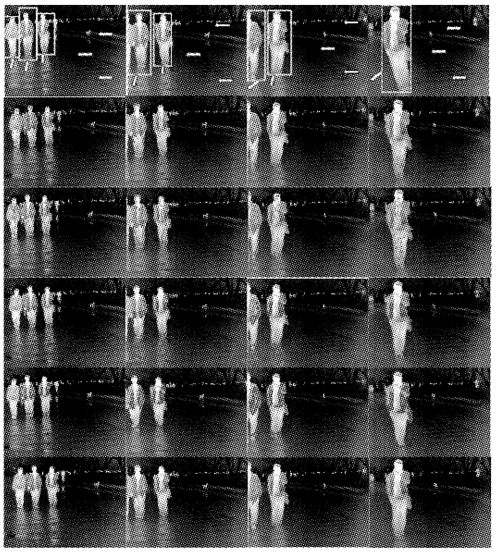


Fig. 10 Segmentation results in clip 3#. First row: Sample frames labeled as $S_1 \sim S_4$ sequentially. Second row: the result of the MOG. Third row: the result by Lee's method. Fourth row: result by Li's method. Fifth row: the result of the PRO. Sixth row: the ground-truth

图10 PRO与各对比方法在clip 3#上的分割结果. 第一行: 依次编号为 $S_1 \sim S_4$ 的序列帧. 第二行: MOG方法结果. 第三行: Lee方法结果. 第四行: Li方法结果. 第五行: PRO方法结果. 第六行: 真实人体区域

to the ground-truth. Let $P = \{p_1, p_2, \cdots, p_n\}$ denote the set of pixels on the boundary, and $Q = \{q_1, q_2, \cdots, q_m\}$ the set of pixels on the ground-truth. The *MAD* is given by

$$MAD(P,Q) = \frac{1}{2}\left\{\frac{1}{n}\sum_{i=1}^{n}\min_{j}||p_i - q_j|| + \frac{1}{m}\sum_{j=1}^{m}\min_{i}||q_j - p_i||\right\}$$

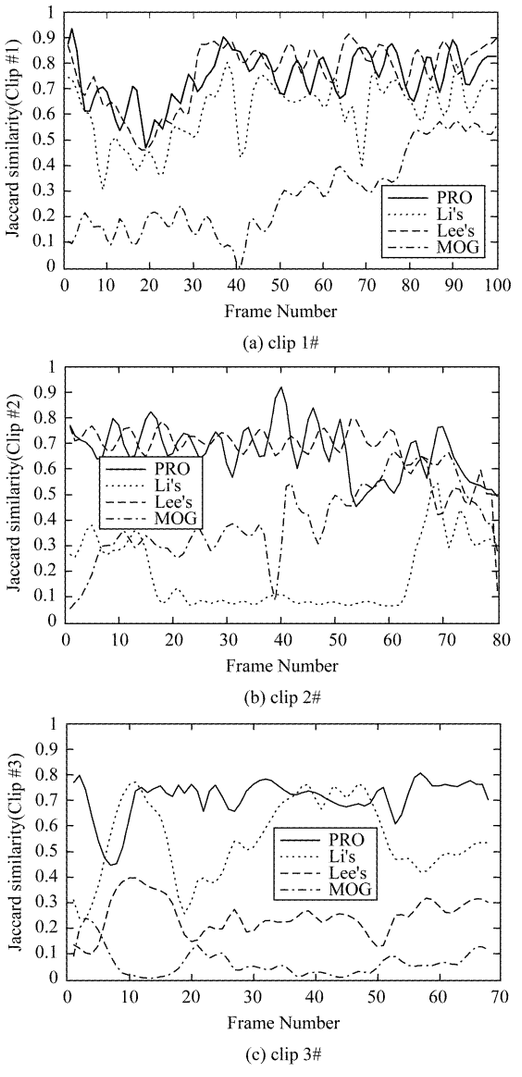


Fig. 11 The comparisons of Jaccard similarity of the MOG, Lee's, Li's and the PRO in test clips

图11 PRO方法与MOG, Lee方法, Li方法在测试序列上的 Jaccard similarity 对比曲线

Figure 11 presents the comparisons of Jaccard similarity of the PRO with its rivals in the test clips. One sees: 1) for the PRO, the averages of the metric S_J. are about 0.8, 0.75, and 0.7 in clip 1# ~ 3#, respectively. Considering poor imagery quality and inhomoge-

Table 2　The threshold values used in the threshold-based LSAC

表 2　基于阈值的 LSAC 模块所使用的阈值

	S_1	S_2	S_3	S_4
clip 1#	$L=120.9$, $U=237.7$	$L=114.8$, $U=232.1$	$L=115.1$, $U=225.2$	$L=116.8$, $U=227.5$
clip 2#	$L=102.5$, $U=237.6$	$L=104.1$, $U=233.8$	$L=97.2$, $U=228.6$	$L=97.5$, $U=230.1$
clip 3#	$L=106.4$, $U=255.0$	$L=107.0$, $U=255.0$	$L=106.7$, $U=255.0$	$L=105.7$, $U=255.0$

subjects, one can believe the PRO based results are good enough; 2) the PRO based curves keep relatively stable while the ones of the rivals fluctuate more violently; 3) the PRO presents the highest (or close to the highest) metric values regardless of the clip captured by a fixed camera or a panning one. However, the metric values of rival methods are sensitive to such factor.

The reasons behind such advantages are: 1) the motion-based LSAC module is dynamic enough to catch possible temporal changes or camera motions owing to the strategy for background update. Since the background is updated by a weighting average of several recent frames, the background image would be close to one of the frames. The background subtraction runs like frame differencing in the module, so it copes well with temporal changes and camera motions; 2) with proper settings of thresholds, the threshold-based LSAC module outputs relatively complete foreground regions, which lead to relatively complete fusion result even if the regions detected by the motion detection module are fragmentary; 3) in the fusion stage, spurious regions can be eliminated. Moreover, human fragments merge to form more meaningful human interiors. The MOG and Lee's methods depend on static background, so that such factors as abrupt intensity changes or camera motion would degrade their performance. Li's method relies on the technique of frame differencing but it leads to many pixel misclassifications.

With refined segmentation results via manually excluding all non-human sub-regions in the foreground, *MAD* computations are conducted and the results are shown in Fig. 12. Since the MOG and Lee's methods fail in clip 3#, corresponding *MAD* curves are not drawn. Figure 12 shows that the *MAD* curves of the PRO keep at a relatively lower level in clip 1# and clip 2#. In clip 3#, the performance of Li's method seems a litter better than the PRO. However, this advantage may be counteracted in practical systems, since the enormous false positives produced may seriously mislead the judgment of human candidates. As a result, the extracted object contour may be completely useless. The advantages of the PRO to present faithful human contour result from the following reasons. Firstly, the LSAC modules themselves present smooth object contours. Secondly, the fusion module never degrades contour smoothness due to its advantage to keep object shapes completely.

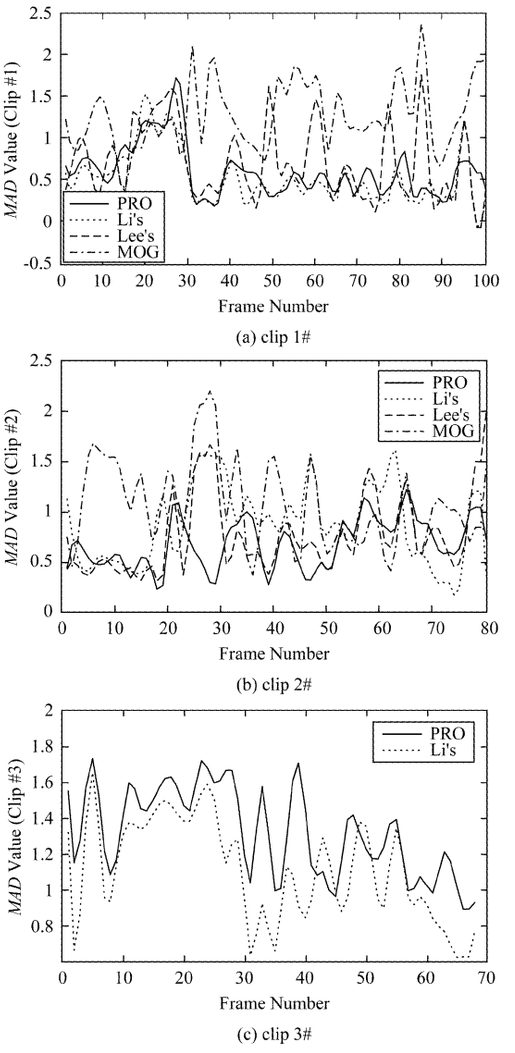


Fig. 12　*MAD* comparisons of the MOG, Lee's, Li's and the PRO in test clips

图 12　PRO 方法与 MOG, Lee 方法, Li 方法在测试序列上的 *MAD* 对比曲线

Table 3 The time costs of level set methods (Sec/frame). (mean ± standard deviation)

表3 各水平集方法在测试序列上的时间开销(秒/帧)(均值±标准差)

	Lee's	Li's	PRO
clip 1#	0.328 ± 0.044	1.293 ± 0.014	0.318 ± 0.027
clip 2#	0.328 ± 0.045	1.333 ± 0.022	0.315 ± 0.005
clip 3#	0.325 ± 0.012	1.388 ± 0.031	0.304 ± 0.005

Table 4 The time costs of the PRO with different parameter settings of the Gaussian filter (Sec/frame). (mean ± standard deviation)

表4 高斯滤波器不同参数值下PRO方法在测试序列上的时间开销(秒/帧)(均值±标准差)

	clip 1#	clip 2#	clip 3#
$K=3,\sigma=0.7$	0.288 ± 0.002	0.296 ± 0.005	0.297 ± 0.006
$K=5,\sigma=1.0$	0.318 ± 0.027	0.315 ± 0.005	0.304 ± 0.005
$K=7,\sigma=1.6$	0.325 ± 0.002	0.325 ± 0.002	0.323 ± 0.002
$K=9,\sigma=2.0$	0.348 ± 0.002	0.347 ± 0.002	0.346 ± 0.002

Next, the computational burden of the PRO is assessed. The average time costs are computed for the methods except the MOG as it is notoriously slow. The results are listed in Table 3. It can be seen that the PRO works as fast as its rivals at least. The reasons behind are: 1) the BLSF based numeric scheme converges so quickly that a solution can be reached after one iteration time at most times. 2) The optimized working flow saves computation costs. For example, this motion-based LSAC module computes only one LSF instead of $(N+1)$ LSFs when it segments a single frame although it requires $(N+1)$ LSFs for the background estimation each time.

The time costs are also recorded for the PRO with different parameter settings of the Gaussian filter. The results are listed in Table 4. As can be seen, the time cost increases with the augment of K. Therefore, relatively smaller K is preferable for efficiency. As for the standard deviation σ, it should be no more than one quarter of K and never be too small for evolution stability. The BLSF is not differentiable at the discontinuities between 1 and -1. The Gaussian filer smoothes the discontinuities and finally results in high quality contours. With a too small σ, the filer may fail to do its job and lead to degraded contours or completely damage the result. For efficiency and accuracy as well, it is appropriate to chose K between $[3, 7]$ and σ between $[0.7, 1.6]$.

4 Conclusion

A level set based algorithm was proposed for the segmentation of infrared human sequences. It provides the results with sub-pixel accuracy, and enjoys the robustness to temporal changes and camera motions. Owing to the fast numeric scheme and optimized working flow, it takes better computational efficiency in comparison with the rivals. Such advantages of the proposed algorithm make it more practical for human detection. However, further algorithmic and code optimizations should be done to meet the real-time demand from practical human detection systems.

REFERENCES

[1] Fernández-Caballero A, Castillo J C, Serrano-Cuerda J, *et al*. Real-time human segmentation in infrared videos[J]. *Expert Systems with Applications*, 2011, **38**(3): 2577-2584.

[2] Li J, Gong W, Li W, *et al*. Robust pedestrian detection in thermal infrared imagery using the wavelet transform[J]. *Infrared Physics & Technology*, 2010, **53**(4): 267-273.

[3] Fernández-Caballero A, Castillo J C, Martínez-Cantos J, *et al*. Optical flow or image subtraction in human detection from infrared camera on mobile robot[J]. *Robotics and Autonomous Systems*, 2010, **58**(12): 1273-1281.

[4] Osher S, Sethian J A. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations[J]. *Journal of computational physics*, 1988, **79**(1): 12-49.

[5] Zhu G, Zhang S, Zeng Q, *et al*. Boundary-based image segmentation using binary level set method[J]. *Optical Engineering*, 2007, **46**(5): 050501-050501-050503.

[6] Vincent L. Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms[J]. *Image Processing, IEEE Transactions on*, 1993, **2**(2): 176-201.

[7] Lee D-S. Effective Gaussian mixture learning for video background subtraction[J]. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2005, **27**(5): 827-832.

[8] Lee S-H, Woo H. Global Illumination Invariant Object Detection With Level Set Based Bimodal Segmentation[J]. *Circuits and Systems for Video Technology, IEEE Transactions on*, 2010, **20**(4): 616-620.

[9] LI Jing, WANG Jun-Zheng, LIANG Shao-Min, *et al*. Method of detecting multiple moving object based on improved level set[J]. *Transactions of Beijing Institute of Technology*(李静, 王军政, 梁少敏,等. 基于改进水平集的多运动目标检测方法. **北京理工大学学报**), 2011, **31**(5): 557-561.

[10] Fang W, Chan K L, Fu S, *et al*. Incorporating temporal information into level set functional for robust ventricular boundary detection from echocardiographic image sequence[J]. *Biomedical Engineering, IEEE Transactions on*, 2008, **55**(11): 2548-2556.