A Local Fuzziness-based Active Contour Model for Infrared Human Segmentation

Yong Tan
School of Electronic & Information Engineering, Yangtze Normal University, Chongqing 408003, China

Abstract
Due to the poor quality of thermal infrared imagery and the great variation of human subjects, segmentation of human subjects is difficult in the thermal infrared images acquired for human detection. In this paper, a local fuzziness-based active contour model is proposed to address this problem. In comparison with such methods as thresholding, it is advantageous to provide more accurate segmentation precision and the enclosed human silhouettes that benefit much incoming human recognition. This model consists of the components named as energy functional and numeric scheme. The energy functional, which decides how a contour evolves over the infrared image domain, is designated from the fuzzy information within local image regions, so that it acquires the ability to differentiate the fuzzy pixels that have similar intensities but actually belong to different regions. The numeric scheme, which evolves the active contour on the image domain by minimizing the energy functional, is carefully formulated via the fusion of the techniques named as direct minimization and narrow band calculation. Solved by this scheme, the minimization of the energy functional is able to reach fast convergence, which ensures the real-time performance of the proposed model. The model is tested on a series of challengeable images that are selected from OSU thermal pedestrian database and the ones acquired by our FLIR A40 thermal infrared camera with the rivals. These images are of different image quality and appearances of human subjects. The results validate its advantages in segmentation accuracy and efficiency, which lead to the adaptation to real-time human detection applications.

Key words: Infrared Detection, Image Segmentation, Active Contour Model, Numeric Scheme.

1. Introduction

Human detection via infrared imagery has been widely studied for such applications as all-weather surveillance, intelligent vehicles and robotic vision. In infrared images, the human subjects of interest are contrasted from the background as long as they have different thermal radiation, so that the infrared imagery dramatically facilitates applications related to human detection. In many human detection applications, image segmentation, which segments the regions of interest that may contain human subjects and would be used for further human/non-human recognition, is critical. Accurate segmentation would ensure the systematic precision, as it dramatically reduces false alarms and provides more details for recognition. However, perfect segmentation, which is expected to detect human subjects with complete interiors and recognizable silhouettes, is hardly available due to several factors. First, infrared images are generally of poor quality, including heavy noise, losses of texture and color, and low-contrast. Second, the variations of human subjects might be great on such aspects as poses, sizes, interior intensities and make human regions hard to be clearly defined. Third, there are many hot objects in surroundings, such as lighting bulbs and working machines. They may degrade or fail the segmentation due to their similar temperature to human bodies.

Various methods have been proposed for infrared human segmentation. Among the methods, thresholding methods [1, 2] might be the most popular base on the assumption that temperature of human bodies is higher than that of environmental objects. However, they often suffer from the problem of fragmentation, which are caused by the un-uniform intensities of human interiors, mutual occlusions, environmental disturbances etc. Generally, remedy operations are necessitated for removal of excessive fragments.

Projection methods [3, 4] segment human subjects by calculating one-dimensional horizontal and vertical profiles of image intensities, because the human subjects are assumed in the intersection regions defined by the thresholded horizontal and vertical image intensity profiles. The methods are generally advantageous in simplicity and efficiency, but they tend to fail to provide meaningful results when meet complex human patterns.

Motion segmentation methods [5, 6, 7, 8] are also popular. They achieve infrared human segmentation by using motion cues of human subjects, and such techniques as background subtraction are frequently used. Unavoidably, precision of these methods relies on heavily the quality of background estimation and they only adapts to the segmentation of human subjects in motion.

As for the fuzzy Clustering methods [8, 9, 10], they assign image pixels to different classes with different memberships instead of assigning them to one class and therefore cope well with the uncertainties of the
segmentation. However, they are fragile to image noises. Frequently, spatial information is incorporated to improve their robustness to image noise.

In comparison, active contour models (ACMs) [11] are more advantageous. First, they have the capability to provide object boundaries with subpixel accuracy. Second, enclosed, smooth contours are inherently available, which facilitates further human recognition. Third, they can be flexibly formulated by integrating various image cues under a relative unified framework. These methods are of the edge-based and the region-based kinds. The edge-based ones detect foreground objects by gradient detection, so that they are unavoidably fragile to noise disturbances and weak boundaries. Unlikely, the region-based ones, such as the Chan-Vese model (C-V) [12], the Fuzzy Energy-based Active Contour (FEAC) [13] or the Local Binary Fitting model (LBF) [14], detect foreground objects by such region cues as intensity averages and probability distributions, and enjoy much better robustness to noise and weak object boundaries. However, as a whole, ACMs are always suffered from computational bottleneck due to the heavy calculation required for contour evolution.

Considering the advantages of ACM methods, a local fuzziness-based active contour model is proposed for infrared human segmentation in this paper. The energy functional incorporates fuzzy logic that copes with the great fuzziness with the infrared image pixels, so that the model has strong ability to differentiate the fuzzy pixels that have similar intensities but actually belong to different regions. Moreover, an innovative numeric scheme, which is formulated based on direct minimization and narrow band calculation, minimizes the energy functional quickly and therefore makes the model satisfy real-time requirements. Experimental results validate its advantages in segmentation accuracy and efficiency. In comparison with the active contour rivals, the proposed model is more applicable for human detection systems that have real-time requirement.

The remainder of this paper is organized as follows. The basic principle of the ACMs is introduced in section 2. Then, the proposed ACM is presented in section 3. Next, experimental results and related analysis are provided in section 4. Finally, some conclusive remarks are given in section 5.

2. ACM Basics

The fundamental idea of ACMs is to start with an active contour around the object of interest. Then, the contour moves toward its interior normal and stops on the object’s boundaries. Let \(C(p,t)\) represent a time-dependent contour that starts from the initial position \(C_0(p)\), and \(\kappa\) is the mean curvature of this contour. The motion of \(C\) can be governed by the following equation

\[
\frac{dC(p,t)}{dt} = F(\kappa)N
\]

(1)

where \(F(\kappa)\) is a function that defines the moving speed of \(C\) along the direction of its Euclidean normal inward vector \(N\). To reach an efficient implementation of (1), a scalar Lipschitz function \(\phi\) called the level set function (LSF) [11] is usually adopted. With the LSF, the contour \(C\) can be implicitly represented as the zero-level set of this function (i.e., \(C = \{p|\phi(p,t) = 0\}\)), and the equation (1) accords with the partial differential equation (PDE) below.

\[
\frac{\partial \phi}{\partial t} = F|\nabla \phi|
\]

(2)

where \(\partial t\) represents the temporal step and \(|\nabla \phi|\) is an appropriate finite difference operator for the spatial derivatives. Governed by (2), the motion of \(C\) is equivalently guided by evolving the LSF \(\phi\), which facilitates the tracking and control of this evolving contour, just as shown in Fig.1.

![Figure 1. The LSF represented active contour and its evolution at different times t.](image-url)
A typical level set-based ACM consists of the components named as the energy functional and the numeric scheme. The energy functional provides a framework that integrates various cues for image segmentation and governs the contour evolution. The numeric scheme solves a minimization problem of the energy functional and gives the segmentation result from the solution it reaches. Generally, the energy functional primarily decides the segmentation accuracy, and the numeric scheme dramatically influences the model efficiency in the meantime affects the numeric accuracy as well.

3. The Proposed Method
3.1. The Fuzziness-Based Energy Functional

Let \( \Omega \) be the domain of the infrared image \( I: \Omega \rightarrow \mathbb{R} \), and \( C \) be the active curve that evolves on \( \Omega \) and separates \( I \) into the interior region (or the object) \( \Omega_1 = \text{inside}(C) \) and the exterior region \( \Omega_2 = \text{outside}(C) \) (or the background). For a given pixel \( x \in \Omega \), the following energy functional can be designated

\[
e(x) = \int_{\Omega_x} u(y)^2 |I(y) - f_1(x)|^2 \, dy + \int_{\Omega_x} |1 - u(y)|^2 |I(y) - f_2(x)|^2 \, dy
\]

(3)

where \( \Omega_x \) is the neighborhood of the pixel \( x \). In this paper, \( \Omega_x \) is chosen as a box-shaped window of size \( w \times w \) for efficiency considerations. \( f_1(x) \) and \( f_2(x) \) are two values that approximate pixel intensities in \( \Omega_1 \) and \( \Omega_2 \), respectively. \( I(y) \) are the intensities in \( \Omega_x \), \( u(y) \), which varies within \([0, 1]\), represents the membership of \( I(y) \) belonging to \( \Omega_1 \). Given a pixel \( x \), the minimum of \( e(x) \) can be reached when the contour \( C \) lies on the object’s boundary. Meanwhile, \( f_1(x) \) and \( f_2(x) \) are optimal to represent the local intensities on the two sides of \( C \). Fig.2 illustrates the definition of the variables in (3).

![Figure 2. Illustration of the variables defined in Eq.(3).](image)

Keeping \( u \) fixed and minimizing \( e(x) \) with respect to \( f_1(x) \) and \( f_2(x) \), respectively, produces

\[
\begin{align*}
f_1(x) &= \frac{\int_{\Omega_x} u(y)^2 I(y) \, dy}{\int_{\Omega_x} u(y)^2 \, dy} = \sum_{y \in \Omega_x} u(y)^2 \frac{I(y)}{\sum_{y \in \Omega_x} u(y)^2} \\
f_2(x) &= \frac{\int_{\Omega_x} |1 - u(y)|^2 I(y) \, dy}{\int_{\Omega_x} |1 - u(y)|^2 \, dy} = \sum_{y \in \Omega_x} |1 - u(y)|^2 \frac{I(y)}{\sum_{y \in \Omega_x} |1 - u(y)|^2}
\end{align*}
\]

(4)

Keeping \( f_1(x) \) and \( f_2(x) \) fixed and minimizing \( e(x) \) with respect to \( u \) produces

\[
u(x) = \frac{(I(x) - f_1(x))^2}{(I(x) - f_1(x))^2 + (I(x) - f_2(x))^2}
\]

(5)

Given an initial partition of the image and initial membership value for \( u \), the object boundary at the position \( x \) can be available by minimizing \( e(x) \). However, it is not enough to derive the entire boundary because \( e(x) \) is locally defined. Instead, it requires the minimization of \( e(x) \) at every pixel in order to attain the entire object boundary. Mathematically, this is equivalent to minimizing the integral of \( e(x) \) over the whole image domain \( \Omega \), namely

\[
E(C) = \int_{\Omega} e(x) \, dx + \mu \oint_C \ell \, ds
\]

(6)

where \( \ell \, ds \) is an arc-length-related regularization term that assures smoothness of the contour, and \( \mu \) determines the weight of this term in the whole integral. When \( E(C) \) has been minimized, the final contour (or the object boundary) and the object region as well can be extracted by
\[
C = \{ x \mid u(x) = 0.5 \}
\]
\[
\Omega = \{ x \mid u(x) > 0.5 \}
\]

The regular approach for minimizing (6) is to derive its Euler-Lagrange equation and to solve it via finite difference iterations. This approach is capable of attaining relatively accurate solution but computationally inefficient due to the short temporal steps that are preferred for iterative stability. To meet the real-time requirement, an alternative scheme that runs much faster is demanded.

3.2. The Fast Numeric Scheme

The regularization term in (6) can be approximated by the sum of the discretized lengths between adjacent points on the contour \( C \) such that
\[
\hat{E}_{\text{C}} = \sum_{i,j} \sqrt{(Q_{i+1,j} - Q_{i,j})^2 + (Q_{i,j+1} - Q_{i,j})^2}
\]
where \( Q_{i,j} = H(u(i, j) - 0.5) \) and \( H(z) \) is a Heaviside function that takes the value of 1 if \( z \geq 0 \) and 0 if \( z < 0 \). Since \( E(C) \) decreases if \( \varepsilon(x) \) decreases at least one pixel, it can be minimized through the following direct minimization.

i. Set an initial partition for the image. For the interior region and exterior region, use \( u > 0.5 \) and \( u < 0.5 \), respectively.

ii. Compute \( f_1(x) \) and \( f_2(x) \) for each pixel in the image domain using (4).

iii. Compute \( \varepsilon(x) \) for each point in the image domain using (3).

iv. Update the membership value \( u(x) \) of the pixel \( x \) only if \( \varepsilon(x) \) decreases, and repeat this operation for each pixel on the image domain.

v. Compute the regularization term using (8) and the total energy \( E(C) \) using (6).

vi. Repeat step 2 to step 5 while the total energy \( E(C) \) keeps decreasing.

This procedure is not as restricted as finite difference schemes and thus it converges fast. However, it suffers from a computational bottleneck since it evolves the contour over the entire image domain. Therefore, a narrow band technique, which restricts the contour evolution in the narrow band region, is used to accelerate the contour evolution process.

![Figure 3. Implicit representation of the image domain.](image)

In ref. [15], \( C \) is represented as the zero-level set of the function \( \phi \) defined as
\[
\phi(x) = \begin{cases} 
3, & \text{if } x \text{ is an exterior point} \\
1, & \text{if } x \in L_{\text{in}} \\
-1, & \text{if } x \in L_{\text{out}} \\
-3, & \text{if } x \text{ is an interior point}
\end{cases}
\]

\( L_{\text{in}} \) and \( L_{\text{out}} \) are two sets of points corresponding to the inner boundary and the outer boundary of \( C \). The definitions of \( L_{\text{in}} \) and \( L_{\text{out}} \) are as follows
\[
L_{\text{in}} = \{ x \mid \phi(x) < 0 \text{ and } \exists y \in N(x) \text{ such that } \phi(y) > 0 \}
\]
\[
L_{\text{out}} = \{ x \mid \phi(x) > 0 \text{ and } \exists y \in N(x) \text{ such that } \phi(y) < 0 \}
\]

where \( N(x) = \{ y \mid x - y = 1 \} \). The points inside \( C \) but not in \( L_{\text{in}} \) are called interior points and those outside \( C \) but not in \( L_{\text{out}} \) are called exterior points. An illustration of such definitions can be seen in fig.3.

The evolution of \( C \) can be represented as the gradient decent flow below
\[
\frac{\partial \phi(x,t)}{\partial t} = \langle F_x + F_y \rangle |\nabla \phi(x,t)|
\]
where $F_i$ is the curvature-dependent smoothing speed. Let $G_\sigma$ be a Gaussian filter with the standard deviation $\sigma$, and let $*$ be a convolution operator. For $\forall x \in L_{in}$,

$$F_i(x) = \begin{cases} -1, & \text{if } G_\sigma * H(-\phi)(x) < 1/2 \\ 0, & \text{otherwise.} \end{cases}$$  \hspace{1cm} (11.1)$$

For $\forall x \in L_{out}$,

$$F_i(x) = \begin{cases} 1, & \text{if } G_\sigma * H(-\phi)(x) > 1/2 \\ 0, & \text{otherwise.} \end{cases}$$  \hspace{1cm} (11.2)$$

$F_d$ is the data-dependent speed, and $\forall x \in L_{in} \cup L_{out}$ here, the speed is set as

$$F_d(x) = u(x) - 0.5$$

with $F_d$ and $F_i$, which respectively correspond to the two terms on the right-hand side of (6), and the contour $C$ evolves to the object boundary using a fast two-cycle process that switches the points of $L_{in}$ and $L_{out}$. By embedding direct minimization into the switching process, a fast numeric scheme can be designated. The detailed procedure of this scheme reads as follows:

\begin{enumerate}
\item \textbf{Initialize $\phi$, $L_{in}$, $L_{out}$, and $u$. For each pixel $x \in \Omega$, set $u(x) > 0.5$ where $\phi(x) < 0$ and $u(x) < 0.5$ where $\phi(x) > 0$.}
\item \textbf{First cycle: data-dependent evolution with the speed $F_d$.}
\begin{enumerate}
\item For $i = 1$: $\text{MaxDeTimes}$
\begin{enumerate}
\item For each $x \in L_{out}$, delete $x$ from $L_{out}$ and add it to $L_{in}$ if $F_d(x) > 0$. Then, add $y$ to $L_{out}$ and set $\phi(y) = 1$ for $\forall y \in N(x)$ with $\phi(y) = 3$.
\item Delete the redundant $x$ from $L_{in}$ if $\forall y \in N(x)$ and $\phi(y) < 0$. Then, set $\phi(x) = 3$.
\item For each $x \in L_{in}$, delete $x$ from $L_{in}$ and add it to $L_{out}$ if $F_d(x) < 0$. Then, add $y$ to $L_{in}$ and set $\phi(y) = -1$ for $\forall y \in N(x)$ with $\phi(y) = -3$.
\item Delete redundant $x$ from $L_{out}$ if $\forall y \in N(x)$ and $\phi(y) > 0$. Then, set $\phi(x) = -3$.
\item For each point in $L_{in}$ and $L_{out}$, update $f_i(x)$ and $f_3(x)$ using (3). Following (5), update the membership value of the point $x$ only if the local energy $\phi(x)$ decreases. Then, update $F_d(x)$ following (12).
\item If the stopping conditions $F_d(x) > 0$ for all $x \in L_{in}$ and $F_d(x) < 0$ for all $x \in L_{out}$ are satisfied, go to the next step. Otherwise, continue this cycle.
\end{enumerate}
\item \textbf{Second cycle: smoothing evolution with the speed $F_i$.}
\begin{enumerate}
\item For $i = 1$: $\text{MaxSeTimes}$
\begin{enumerate}
\item For each $x \in L_{out}$, delete $x$ from $L_{out}$ and add it to $L_{in}$ if $F_i(x) > 0$. Then, add $y$ to $L_{out}$ and set $\phi(y) = 1$ for $\forall y \in N(x)$ with $\phi(y) = 3$.
\item Delete the redundant $x$ from $L_{in}$ if $\forall y \in N(x)$ and $\phi(y) < 0$. Then, set $\phi(x) = 3$.
\item For each $x \in L_{in}$, delete $x$ from $L_{in}$ and add it to $L_{out}$ if $F_i(x) < 0$. Then, add $y$ to $L_{in}$ and set $\phi(y) = -1$ for $\forall y \in N(x)$ with $\phi(y) = -3$.
\item Delete redundant $x$ from $L_{out}$ if $\forall y \in N(x)$ and $\phi(y) > 0$. Then, set $\phi(x) = -3$.
\end{enumerate}
\item Output the contour $C = \{ x \mid \phi(x) = 0 \}$ and the object region $\Omega = \{ x \mid \phi(x) < 0 \}$ if the stopping conditions $F_i(x) \geq 0$ for all $x \in L_{in}$ and $F_i(x) < 0$ for all $x \in L_{out}$ are satisfied, or go to step 2 if not.
\end{enumerate}
\end{enumerate}
\end{enumerate}

One should note that under a situation with a fixed Gaussian filter $G_\sigma$, the parameter $\mu$ in (6) (i.e., the regularization strength) is controlled by the parameter $\text{MaxDeTimes}$ and $\text{MaxSeTimes}$, and the larger $\text{MaxSeTimes}$ is relative to $\text{MaxDeTimes}$, the more strength it provides.

\textbf{4. Experimental Results and Analysis}

The proposed model (PRO) has been implemented by MATLAB 2016a on an ASUS laptop with an Intel Core i7 4720HQ 2.60 GHz CPU, 4G RAM, and the Windows 10 operating system. In this section, infrared human images are selected from the “OSU Thermal Pedestrian Database” [16] and custom databases produced by our FLIR A40 thermal infrared camera. PRO is parameterized as follows: $\text{MaxDeTimes}=30$, $\text{MaxSeTimes}=1$, $\sigma=1$, $w=5$, the initial membership value $u$ is 0.9 for the interior region and 0.1 for the exterior region. For comparison, the classic C-V [12] is chosen as the representative of ACMs with the assumption of global intensity homogeneity, the FEAC [13] as the representative with global fuzzy information, and the LBF [14] as the representative dealing with heterogeneity by use of local intensity information. To attain as better results as possible, all parameters that rival methods use are tunable. Besides, the initial partitions are identically set for PRO and the rivals as well.
In this first test, three representative images shown in the top row of figure 4 are adopted. These images are of 320×240 resolutions. They are characterized by heavy noise, low-contrast and the human subjects having burry boundaries and heterogeneous interiors. It is challengeable for the referred methods to reach successful segmentation.

The PRO and the rivals are applied to these images, and the results are respectively given in the left rows of figure 4. With reference to the ground truths represented by red contours drawn on the test images in the top row, one sees that the C-V misclassifies lots of background pixels and fails to locate human boundaries. This failure can be attributed to the statistical homogeneity assumption that fails to differentiate the foreground and background regions having very close intensity averages, especially under the absence of gradient detection.

![Segmentation results given by PRO and the rivals.](image)

**Figure 4.** Segmentation results given by PRO and the rivals.

Unlikely, the LBF locates exact human boundaries, which validates its advantage in differentiating the regions featured by close intensities, but the fragmentation in human interiors makes meaningful subject of
interest unavailable. The FEAC evolves the contours based on global fuzziness instead of intensity averages, it produces fewer misclassification errors of background pixels when compared with the C-V and the LBF due to its good capability to reject weak local minima. However, as the membership values are not updated in a limited region around the contour but over the whole image domain, contour discontinuities may occur and therefore cause a failure in producing exact object boundaries. Clearly, the PRO provides the results that are most faithful to ground truths. The reasons behind can be given as follows. First, fuzzy logic is effective in dealing with the similarity between the pixels belonging to different regions. Second, the participation of local information enables PRO to address weak human boundaries and heterogeneous human interiors. Third, by conducting the contour evolving in a narrow band region, the update of membership values is only conducted at the pixels on the contour (i.e., the pixels in \( L_{ac} \) and \( L_{ab} \)), so that the PRO avoids the occurrence of contour discontinuities.

Moreover, efficiency of these methods is compared by calculating the average CPU time cost of three trials of segmenting each image, and the results have been listed in Tab.1. Both the C-V and LBF run very slowly, because they evolve active contours on the whole image domain and use the traditional difference scheme that slowly converges. The FEAC runs much faster by using the direct minimization scheme, which converges after only one sweep on the whole image domain. By contrast, the PRO costs the least time, which validates the efficiency of the proposed numeric scheme.

The efficiency of PRO is further explored under the situation with different parameter settings. Varying the parameter \( MaxSeTimes \) from 1 to 30 and recording the average time costs, figure 5 can be drawn. As seen, the time costs linearly increase as \( MaxSeTimes \) increases, which embodies influence of the regularization strength on convergence speed. Additionally, the costs that change with the parameter \( w \) are recorded and then listed in Tab.2. One sees that this parameter \( w \) does affect the efficiency although it is far less dramatic than \( MaxSeTimes \). There are two reasons behind. First, the number of iterations before numeric convergence would not change with \( w \) due to the effectiveness of fuzzy logic. Second, the points whose membership values need to be updated are rigidly limited in two lists. To sum up, the relatively small regularization strength and window sizes are more preferable for PRO.

### Table 1. Comparison of the average CPU time costs of the ACMs (Seconds per trial).

<table>
<thead>
<tr>
<th>Image No.</th>
<th>C-V</th>
<th>LBF</th>
<th>FEAC</th>
<th>PRO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image #1</td>
<td>0.684286</td>
<td>10.416851</td>
<td>0.070562</td>
<td>0.000424</td>
</tr>
<tr>
<td>Image #2</td>
<td>0.409067</td>
<td>2.116659</td>
<td>0.061029</td>
<td>0.000358</td>
</tr>
<tr>
<td>Image #3</td>
<td>0.155645</td>
<td>3.163194</td>
<td>0.051076</td>
<td>0.000471</td>
</tr>
</tbody>
</table>

### Table 2. Time costs of the PRO parameterized by different window sizes. (Seconds per trial).

<table>
<thead>
<tr>
<th>( w )</th>
<th>Image #1</th>
<th>Image #2</th>
<th>Image #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( w = 7 )</td>
<td>0.000460</td>
<td>0.000461</td>
<td>0.000487</td>
</tr>
<tr>
<td>( w = 9 )</td>
<td>0.000588</td>
<td>0.000558</td>
<td>0.000600</td>
</tr>
<tr>
<td>( w = 15 )</td>
<td>0.000777</td>
<td>0.000590</td>
<td>0.000637</td>
</tr>
</tbody>
</table>

Again, PRO is tested under the context of human detection. In this test, PRO and the rivals are initially applied to segment warmer blobs that may contain human subjects from test sequences. Then, the criteria, including the area, area ratio, dispersion and histogram distribution, are applied for recognizing the blobs. At last, to calculate a series of indices including false positives (FP), false negatives (FN), true positives (TP), precision and Jaccard coefficient [17], it reaches the objective evaluation of these methods. Note that FP, FN and TP mean the incorrectly classified foreground objects, the incorrectly classified background objects and the correctly classified foreground objects, respectively. The definition of precision is \( PR = TP / (TP + FP) \) and it reflects the ratio of correctly classified objects to total objects. As for the Jaccard coefficient, it is defined by \( S_j = TP / (TP + FP + FN) \). This index hits 0 when there are no matching blobs and 1 when a perfect match occurs.

Three sequences selected from the OSU Thermal Pedestrian Database and our custom ones are used for this test. The sequences differ in human sizes, noise intensity, boundaries and heterogeneity conditions. More details of these sequences are provided in tab.3.
Table 3. Details of the test sequences.

<table>
<thead>
<tr>
<th>Sequence No. Characteristics</th>
<th>Sequence #1</th>
<th>Sequence #2</th>
<th>Sequence #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample frame</td>
<td><img src="https://example.com" alt="Image" /></td>
<td><img src="https://example.com" alt="Image" /></td>
<td><img src="https://example.com" alt="Image" /></td>
</tr>
<tr>
<td>Originality</td>
<td>From the OSU Thermal Pedestrian Database</td>
<td>Shot by fixed FLIR A40 camera.</td>
<td>Shot by fixed FLIR A40 camera.</td>
</tr>
<tr>
<td>Frame number</td>
<td>23</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Noise level</td>
<td>Fair</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Human sizes</td>
<td>Small</td>
<td>Middle/Large</td>
<td>Middle/Large</td>
</tr>
<tr>
<td>Contrast</td>
<td>Low</td>
<td>Good</td>
<td>Low</td>
</tr>
<tr>
<td>Boundary</td>
<td>Highly blurred</td>
<td>Blurred</td>
<td>Highly blurred</td>
</tr>
<tr>
<td>Inhomogeneity</td>
<td>High</td>
<td>Fair</td>
<td>High</td>
</tr>
</tbody>
</table>

Table 4. Objective evaluations of the PRO and its rivals in human recognition test.

<table>
<thead>
<tr>
<th>Method</th>
<th>C-V</th>
<th>LBF</th>
<th>FEAC</th>
<th>PRO</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>92</td>
<td>87</td>
<td>83</td>
<td>91</td>
</tr>
<tr>
<td>FP</td>
<td>3</td>
<td>8</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>FN</td>
<td>45</td>
<td>23</td>
<td>23</td>
<td>7</td>
</tr>
<tr>
<td>(PR)</td>
<td>0.9684</td>
<td>0.9158</td>
<td>0.8737</td>
<td>0.9579</td>
</tr>
<tr>
<td>(S_J)</td>
<td>0.6571</td>
<td>0.7373</td>
<td>0.7034</td>
<td>0.8922</td>
</tr>
<tr>
<td>TP</td>
<td>189</td>
<td>192</td>
<td>190</td>
<td>194</td>
</tr>
<tr>
<td>FP</td>
<td>11</td>
<td>8</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>FN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(PR)</td>
<td>0.9450</td>
<td>0.9600</td>
<td>0.9500</td>
<td>0.9700</td>
</tr>
<tr>
<td>(S_J)</td>
<td>0.9450</td>
<td>0.9600</td>
<td>0.9500</td>
<td>0.9700</td>
</tr>
<tr>
<td>TP</td>
<td>284</td>
<td>298</td>
<td>281</td>
<td>318</td>
</tr>
<tr>
<td>FP</td>
<td>50</td>
<td>26</td>
<td>43</td>
<td>6</td>
</tr>
<tr>
<td>FN</td>
<td>29</td>
<td>11</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>(PR)</td>
<td>0.8503</td>
<td>0.9198</td>
<td>0.8673</td>
<td>0.9815</td>
</tr>
<tr>
<td>(S_J)</td>
<td>0.7824</td>
<td>0.8896</td>
<td>0.8673</td>
<td>0.9008</td>
</tr>
</tbody>
</table>

Some representative results produced by PRO are given in fig.6. In this figure, the segmented human subjects are denoted by green bounding-boxes and their boundaries are denoted by red contours. As observed, both false positives and false negatives are rare, and the contours fit exactly human boundaries. By contrast, the rival-based results shown in fig.7 have more segmentation errors. From tab.4, in which the objective evaluation results are listed, one sees the dramatic advantage of PRO in segmentation accuracy especially when the sequences are of poor quality. It is reasonable to believe the PRO would lead to good recognition precision in human detection applications.
(a) The results on sequence #1

(b) The results on sequence #2

(c) The results on sequence #3

Figure 6. Segmentation results given by the PRO.

(a) C-V based results

(b) LBF based results

(c) FAC based results

Figure 7. Segmentation results on sequence #1 given by the rival methods.
5. Conclusions

In this paper, a local fuzziness-based active contour model is proposed for infrared human image segmentation. This model is advantageous in segmentation accuracy and efficiency due to the energy functional and the fast numeric scheme. The energy functional that has been formulated with fuzzy logic copes well with pixels that have similar intensities but actually belong to different regions, and the numeric scheme that has minimizes the energy functional has real-time efficiency, which makes the model adaptive to time demanding human detection applications. In future, further algorithmic optimization and implementation on specific hardware platforms should be done for developing real-time infrared human detection systems.

Acknowledgements

This work was partially supported by Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant No. KJ1601204) and the Initial Funding of Scientific Research for Introduction of Talent of Yangtze Normal University (Grant No.2014KYQD07).

References


