Automatic Target Segmentation by Locally Adaptive Image Thresholding

Wen-Nung Lie

Abstract—A locally adaptive thresholding algorithm, concerning the extraction of targets from a given field of view, is proposed. Conventional histogram-based or global-type methods are deficient in detecting small targets of possibly low contrast as well. Our research is notable for solving the mentioned problems by introducing 1) shape connectivity measure based on co-occurrence statistics for threshold evaluation; and 2) no-target identification procedure for modeling a local-processing paradigm. In this manner, thresholds are determined adaptively even in the presence of space-varying noise or clutters. Experiments show that our results are reliable and even outperform those that manual operations can achieve for global thresholding.

I. INTRODUCTION

The problem of segmenting an object from a given field of view is basic to many image processing applications. Examples of such objects include lung tumors in chest radiographs, nuclei of blood cells and chromosomes in microscope images, solids in range images, or even military targets in infrared and visible images. The digitization of such images in image space might have a poor target/background contrast and unwanted background clutters may make segmentation difficult.

The spoke filter [1] detects blobs of convex shapes by incorporating the edge information. Its demand in tuning parameters disables itself from reaching an automatic operation. Duda and [2] adopted a two-step process where a size-contrast filter of double gates roughly locates the areas of interest, followed by an intensity-gradient approach to extracting objects. This method requires accurate knowledge of target size to maintain least errors. Other approaches include model-based feature groupers [3] and multi-resolution pyramid method [4].

The most popular approaches to target segmentation may be the thresholding techniques [5]. Traditional algorithms considered a fixed threshold value according to gray-level histogram and cannot process images whose histograms are nearly unimodal, especially when the target region is much smaller and low-contrasted relative to the background area. Dunn et al. [6] adopted "uniform error thresholding" to be capable of segmenting small objects, but restrictedly assumed uniform statistics in both the target and background regions. The above methods are commonly lacking of a measure of thresholding quality and consequently, on-line rejection of poor results and reduction of false alarms seem difficult [7], [8].

Our algorithm is in contrast locally adaptive, getting multiple or region-dependent thresholds based upon a distribution-free local analysis of the image. In prior literature, Nakagawa et al. [9] divided an image into small windows whose thresholds are then interpolated for the entire image. Their method, however, requires the bimodality of each window. Parker [10], on the other hand, noticed and coped with the problem of thresholding in bad illumination by locating object pixels first (using local intensity gradient feature), growing regions around them, and then generating a threshold value for each pixel. Our method—applying a scanning window to obtain local thresholds for combination—is specific in three aspects.

1) Local thresholding is performed based on the second-order co-occurrence statistics [11], [12], rather than on conventional graylevel histogram [13]-[15]. This kind of principle was found effective in target segmentation applications [11], [12], [16].

2) Target/background a priori information is used to help the localization of small targets.

3) Region-dependent thresholds result in a superior adaptivity to space-varying backgrounds, hence requiring less post-processing for removing superfluous clutters.

II. A NEW MEASURE BASED ON CO-OCCURRENCE MATRIX-SHAPE CONNECTIVITY

Following Chanda and Majumber [12], [18], we define g(j, k) to be an L-level graytone image of size M × N and its co-occurrence matrix as

\[ [C] = \{ c_{m,n} | 0 \leq m, n \leq L - 1 \} \]

Letting \( t \) be the level for image thresholding, it partitions the co-occurrence matrix into four distinct blocks, \( B_1(t), B_2(t), B_3(t) \), and \( B_4(t) \), as shown in Fig. 1. The computing formulas of them can be seen in [12], [18].

We now propose a "shape connectivity" (SC) measure for which the optimal threshold is chosen as

\[ t^* = \arg \left[ \text{Maximize} \left( \frac{\text{Minimum} \left( B_1(t), B_2(t) \right)}{B_3(t) + B_4(t)} \right) \right] \]

in which \( w_{m,n} = 1 \) for \( 0 \leq m, n \leq L - 1 \) (see [12], [18]).

According to the basic properties of \( B_i(t) \) [12], [18], our SC measure, the bracketed item in (1), meaningfully represents the area/perimeter ratio of the minor class region. Fig. 2 gives simple examples illustrating binary images of different SC values, where shaded areas represent the object pixels.

Evidently, we may have distinct SC values for the same object areas, while in support of compact regions against noisy ones. Furthermore, the SC measure emphasizes compact regions of larger areas.

Fig. 1. Partition of the co-occurrence matrix \( \{ c_{m,n} \} : B_1, B_2, B_3, B_4 \).

Fig. 2. Binary images of different shape connectivities. (a) SC(t1) = Min (8, 24)/16 = 1/2, (b) SC(t2) = Min(0, 21)/26 = 0.

Authorized licensed use limited to: Shantou University. Downloaded on December 30, 2008 at 10:42 from IEEE Xplore. Restrictions apply.
Authorized licensed use limited to: Shantou University. Downloaded on December 30, 2008 at 10:42 from IEEE Xplore. Restrictions apply.
compute $SC(t)$ for $t=0, ..., L-1$.

Smooth $SC(t)$ curve with kernel $[p, l-2p, p]$ when $0 < p < 1$.

Find $t_1$.

Locate peak $t_2$ from interval $(t_c, L)$ for which $SC(t_2) > SC(t_1 + i)$, for $i = -2, ..., 2$.

If $t_2$ exists, yes.

Otherwise, NO.

Select $t_m = t_1$ or $t_2$, which gets a larger $SC$ value.

If $t_1$ exists, yes.

Otherwise, NO.

Select $t = L$ (no target).

$t^* = L$ (no target).

Fig. 7. Algorithm of image thresholding with no-target identification.

C. Determination of Locally Adaptive Thresholds

Denote $T = \{t_m | m = 0, \ldots, (M-H)/\Delta y, n = 0, \ldots, (N-W)/\Delta x\}$ to be local thresholds obtained by the above mentioned paradigm. For each pixel $g(j,k)$, we can find $D^{jk} = \{t_{il} | h = n_i, i = n_i, l = n_i, \ldots, n_i\}$, a submatrix of $T$, to which $g(j,k)$ has contributions (i.e., $g(j,k)$ is contained in $t_{il}$'s corresponding window). Our criterion is thus to determine

$$g(j,k) := \text{background, if} \sum_{h,l} \#(g(j,k) < t_{hl}) > \sum_{h,l} \#(g(j,k) \geq t_{hl})$$

$$g(j,k) := \text{target, if} \sum_{h,l} \#(g(j,k) < t_{hl}) \leq \sum_{h,l} \#(g(j,k) \geq t_{hl})$$

Notice the requirement of "peak.thd" to reject patterned clutters that cause small peaks in $SC(t)$ curve. Our algorithm is also promising for segmentation of targets with contrast-reversal (e.g., targets colder than the background in thermal imageries). This simply leads to the refinement in Fig. 8.

Fig. 8. Refined algorithm for thresholding on contrast-reversed images.
Fig. 9. Incremental co-occurrence matrix computation. Voting from the central shaded area remains unchanged, while that of the other two distinct regions contributes positively and negatively, respectively.

\[ \sum_{h,l} \#(g(j,k) \geq t_{sl}) \geq \sum_{h,l} \#(g(j,k) < t_{sl}) \]  

(9)

where "\#" represents "is classified as" and "\# (event)" equals 1 if "event" is true and 0 otherwise. That is, we count the classification, by thresholds \( t_{sl} \)'s, as "target" and "background," and adopt the dominant one as final decision. Actually, this leads to the following equivalence.

**Property 4:** The same classification can be achieved by determining an equivalent threshold \( \bar{t} \)

\[ \bar{t} = D'[q], \quad q = \lceil WH/(2\Delta x \cdot \Delta y) \rceil + 1 \]  

(10)

where \( D' \) is a sorted sequence, in an increasing order, of elements of \( D = \{t_{sl}\} \).

From (10), \( \bar{t} \) is obviously the median of the \( D' \) sequence. The classification by \( \bar{t} \) is the same as that by direct comparisons (8), (9)), but with higher computational complexity (due to sorting).

IV. COMPUTATIONAL EFFICIENCY

According to our computing flow, computations of \( \{c_{m,r}\} \) and \( SC(t) \) curve at varying positions of local windows, are by far the dominating sources for CPU time.

A. Efficient Computation of \( B_1(t) \rightarrow B_2(t) \)

The derivation of \( SC(t) \) is heavily dependent on the cooccurrence voting and computation of \( B_1(t) \rightarrow B_2(t) \). We have proposed in our prior research [18] an efficient algorithm that directs the voting onto two 1-D arrays, \( \{b_i\} \) and \( \{\sigma_i\} \), instead of on the 2-D array, \( [C] \).

B. Incremental Co-Occurrence Matrix Computation

Fig. 9 illustrates two overlapping windows displaced by \( \Delta x \) in the horizontal direction. Let their co-occurrence matrices be \( [C]_l \) and \( [C]_r \), respectively. As the voting from the overlapping area remains unchanged, \( [C]_r \) can be described as

\[ [C]_r = [C]_l - [C]_{lr} + [C]_{rp} \]  

(11)

where \([C]_{lr}\) stands for contributions from the left portion to be subtracted (negative) and \([C]_{rp}\) stands for contributions from the right area which are newly formed (positive). Accordingly, totally \( 2 \cdot \Delta x \cdot H \) times (instead of \( W \cdot H \)) of voting are performed at each position of the sliding window.

<table>
<thead>
<tr>
<th>TIME PER OPERATION</th>
<th>( O_t )</th>
<th>( O_{op} )</th>
<th>( O_{op} )</th>
<th>( O_d )</th>
<th>( O_{op} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total operations</td>
<td>( 2H \Delta x \Phi )</td>
<td>( 4L \Delta x \Phi )</td>
<td>( H \Phi )</td>
<td>( W \Phi )</td>
<td></td>
</tr>
<tr>
<td>* Time for INT_ADD, INT_MUL, FLOAT_ADD, and FLOAT_MUL are estimated to be ( k \cdot 1.5 \Delta x ) ( , 2 \Delta x ) ( , 2 \Delta x ) ( , ) respectively. * Denote ( \phi = \Phi((\Delta x, \Delta x)) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE II

TIMES AND NUMBERS FOR FIVE CATEGORIZED OPERATIONS IN OUR LOCAL THRESHOLING ALGORITHM

C. Time Complexity

Arithmetic operations involved in our local thresholding algorithm can be categorized as follows.

1) Voting operations on the proposed 1-D \( \{b_i\} \) and \( \{\sigma_i\} \) arrays [18], denoted as \( O_{vt} \). For most pixels, except those of the rightmost and bottommost rectangular areas, voting is performed \( 2H/\Delta y \) times (negative and positive for \( H/\Delta y \) horizontal scans); thus totaling more than \( 2H/\Delta y \cdot (M - H)/(N - W) \) operations for the whole processed area.

2) Incremental summation, denoted as \( O_{op} \), for computing \( B_1 \), \( B_2 \) and \( B_3 \) [18]. It totals an amount of \( 4L(\Delta x \cdot M - H)/\Delta y \cdot (N - W)/\Delta x \) (18) for the whole image area.

3) Operations for computing \( SC \), denoted as \( O_{op} \), total an amount of \( L(\Delta x \cdot M - H)/\Delta y \cdot (N - W)/\Delta x \).

4) Operations for smoothing \( SC(t) \), denoted as \( O_{sm} \), total an amount of \( L(\Delta x \cdot M - H)/\Delta y \cdot (N - W)/\Delta x \).

5) Counting of classification (8), (9), denoted as \( O_{sc} \), approximates to \( MN \cdot (W/\Delta x \cdot H/\Delta y) \).

Fig. 10. Experiments on a synthetic image of space-varying statistics. (a) Original image; (b) Otsu's method, \( t^* = 126, 0.174 \) s; (c) moment-preserving method, \( t^* = 138.0.188 \) s; (d) CP measure (global), \( t^* = 167.0.367 \) s; (e) SC measure (global), \( t^* = 167.0.366 \) s; (f) CP measure (local, \( W = H = 35, \Delta x = \Delta y = 1, t_v = 6 \), 203.01 s; (g) SC measure (local, \( W = H = 35, \Delta x = \Delta y = 1, \) peak_thd = 0.7), 180.14 s; (h) Parker's method, gradient_thd = 60.623 s.
We note a time complexity of $O(L^2)$, i.e., $O(LMN)$, in Table II that summarizes the above analysis. Formally, all of the five categories of operations can be hardware-implemented and pipeline-executed, in view of their regularity and independence.

V. EXPERIMENTS

First, we experiment with a synthetic image containing four rectangles (Fig. 10). The background statistics is made space-varying by increasing the Gaussian mean horizontally, but keeping a constant variance. Rectangle graylevels are also of normal distributions. To illustrate the superiority of the proposed algorithm, we compare several global and local thresholding techniques: b) Otsu's method [13]; c) Tsai's moment-preserving method [14]; d) conditional probability (CP) measure (global), $t^* = 2, 0.18 s$; (f) SC measure (global), $t^* = 256, 0.091 s$; (e) CP measure (local, $W = H = 40, \Delta x = 1, t_r = 6$), 3.521 s; (b) SC measure (local, $W = H = 40, \Delta x = 1, \text{peak.thd}=0.7$), 3.406 s; (i) Parker's method, gradient.thd = 10, 2.945 s.

As expected, Fig. 10(b) and (c) justify that histogram-based methods are deficient in the capability of segmenting small objects. Conditional probability and shape connectivity measures implemented in a similar situation of locally adaptive versions, as stated in Section III, behave differently (Fig. 10(f) and (g)) in spite of the same results obtained in their global manner (Fig. 10(d) and (e)). To establish no-target identification for the CP measure, we make a conjecture that

$$|\mu - \mu| \leq t_r,$$

where $\mu$ is the image mean and $t_r$ represents an operational tolerance. It is evident from Fig. 10(f) that condition (12), with $t_r = 6$, does not lead to a satisfactory result. Only small improvement is achieved even if $t_r$ is increased up to 20 (the result is not shown here). On the other hand, the benefit of SC measure as well as its locally adaptive implementation can be proved in view of Fig. 10(e) and (g). We also choose for comparison Parker's algorithm [10] (Fig. 10(h)), which was developed to solve the problem of bad or nonuniform illumination. His method seems to fail in object aggregation for such noisy images.

Then we demonstrate with real images. Figs. 11–13 show forward-looking infrared (FLIR) images which have been preprocessed such that targets are located to reside in a horizontal zone. Similarly, each experiment is performed using the above-listed seven methods ((c)–(i)), plus subjectively manual thresholding ((b)). Evidently no global methods, including the manual type that generate one single threshold, can outperform our local version.

It is noted that Fig. 11(f) falsely produced a full background classification. This result is important and actually more preferable than Figs. 11(c)–(e) in military applications where false alarms should be kept at as low a constant as possible.
Parker's algorithm seems to offer a compromise between performance and speed. However, its requirement of a prefixed gradient threshold in order to locate initial object pixels presents us another thresholding problem. Parameter "gradient of threshold" is unfortunately likely to vary case-by-case (Figs. 10–13).

As for the CPU time (based on SUN 4/Sparc 2 workstation), it can be seen from captions of Figs. 11–13 that locally adaptive methods generally spend far more time (1–2 orders, depending on 1-D or 2-D scanning) than the global ones, as analyzed previously. To speed up, either $\Delta x$ and $\Delta y$ should be enlarged or VLSI hardware finally be used.

VI. CONCLUDING REMARKS

We have proposed a local thresholding paradigm for segmenting small and low-contrast targets out of a noisy background. Success of this paradigm is crucially contributed from incorporating no-target identification in local areas, which basically forms a significant difference from conventional algorithms. Actually, it is possible to combine other global techniques into our local paradigm with the association of their own no-target identification procedures (e.g., Fig. 10(f)).

Currently, our implementation is limited to (1) binary thresholding, and (2) random background noise. For multi-level thresholding, a more sophisticated identification of modes $t_{p1}, t_{p2}, t_{p3}$, . . . from noisy $SC(t)$ curve will be required. On the other hand, patterned noise possibly makes no-target identification ineffective and causes false alarms or processing errors.

Our local method is promising in three respects: 1) adaptivity to space-varying statistics; 2) achieving the most human perception by considering explicit features (i.e., shape connectivity) of the thresholded patterns; and 3) being real-time promising due to efficient computations implementable using ASIC's and high-speed DSP chips.

REFERENCES