OBJECT RECOGNITION THROUGH TEMPLATE MATCHING USING AN ADAPTIVE AND ROBUST HAUSDORFF DISTANCE

Yong Hu 1, Weiling Xia 2, Xiangyun Hu 1, Ruisheng Wang 1
1 Department of Earth and Space Science, York University
2 Shenzhen Bureau of Municipal Construction Archives
E-mails: {yhu; xyhu; rwang}@yorku.ca; mollyxia@hotmail.com

ABSTRACT

In computer vision, the partial Hausdorff distances (PHDs) are used to compare images but with strong limitations of using fixed fractions. In this paper, we formulate an adaptive and robust Hausdorff distance (ARHD) with non-parametric and robust statistical methods. The new distance is estimated using the empirical distribution of the distance variable based on the distance map of the template’s edge map, and this makes full use of the information associated with the edge distribution structure of templates. The best fraction is determined by adaptively adjusting at two directions along the distance curve and the statistical test using linear regression. The distance threshold is also derived from the same empirical distribution function estimated from the template. Therefore, it is not sensitive to the initial fraction values compared with conventional PHDs. The experiments using aerial images show that ARHD has good performance in matching templates in heavily blurred and complex backgrounds with pose change, scale change, geometric distortion and partial occlusion.

INTRODUCTION

For an image, what people are most interested in are some dominate objects. Object recognition is to identify and locate corresponding points or regions by image matching methods in two or multiple images collected at a same scene. It is one of the key topics in computer vision and digital photogrammetry. For the tasks of machine vision, pattern recognition and digital surface model generation, we have to determine if some prototype or template exists in an image, and get the best registration transformation that describes their relative attitudes. This is to find a transformation between the template and the reference image so that each pixel in the template corresponds to an appropriate pixel in the reference image.

To recognize those objects, object features are usually extracted. Most matching methods used to solve this type of tasks can be classified into three classes (Borgerfors, 1988): (a) Algorithms using raw data such as pixel gray values. For example, the correlation technique in space domain uses the correlation functions to identify corresponding points or regions. And the Fourier-Mellin transformation searches peaks of the phase correlation in frequency domain (Jian et al., 1999); (b) Algorithms using low-level features. For example, point-set matching techniques use interest points, corners, edges and lines; (c) Algorithms using high-level features such as the relations between low-level features in relation matching. The first method is based on the correlation measure between images, is sensitive to noises, and has great computational burden. Multiple peaks will occurs when there are narrow random noises with high frequencies in the image, and large false alarms will occurs at occluded areas and areas shortage of features. While the third method requires extracting and labeling the relations among low-level features, and it itself is a difficult matching task. The second method is expected to output better results due to the fact that low-level features can express some stable structures and are usually more reliable than the pixel values.

At present, the point-set matching techniques have become a popular image matching method. It can be described as: given two images and the allowable transformation space $T$ that transform one image to another one, two point sets $A$ and $B$ detected from two images, and the similarity measure between point sets. We have to search for such a transformation $t \in T$ so that the distance between $t(B)$ and its corresponding object in $A$ (called search map $A_t$) is minimized. A transformation $t \in T$ maps a pixel of $B$ to a point in the coordinate system of $A$. The mapped image block of $B$ is denoted as $t(B)$. The region covered by $t(B)$ in $A$ is called the search map and is denoted as $A_t$. While the inverse transformation $t^{-1}$ maps every pixel of $A_t$ to a point in the coordinate system associated with $B$, and the image block produced is denoted as $t^{-1}(A_t)$.

Obviously, the attitude and the transformation parameters can be computed correctly using a small number of conjugates and the transformed template would be exactly same with its image conjugate if there are no non-
deterministic disturbances in sensing and no wrong conjugate features between the template and the reference image. However, there is great illumination difference between the template and the image because of different radiation distortion, geometric distortion, random noise when imaging at different time and conditions. These noises will be propagated to attitude parameters. In addition, the extent affected by the indeterminate disturbs relies on the features used to estimate the attitude parameters. This results in the difficulty in forming a robust similarity measure. Therefore, it is important to use a similarity measure to some extent it can resist disturbs. The gray correlation function, absolute difference, single point least-squares matching and recognition based on features and models establish the point-to-point corresponding relationship between the features of model and reference image. These types of algorithms can obtain a high accuracy, but often result in high computational complexity (Ackerman, 1984). Another type of algorithms needs no explicit point-to-point corresponding (Borgerfors, 1988; Huttenlocher et al., 1993; Rucklidge, 1997; Jian et al., 1998). A famous one is the partial Hausdorff distance proposed by Huttenlocher et al. (1993), which uses two fixed fractions of all the distance values to detect a part of templates. Thereafter, several modifications to the PHD were proposed (Kwon, 1996; Olson, 1997; Hu et al., 1999). However, aforementioned Hausdorff distances assume some special requirements, and are valid only for certain errors. When multiple errors occur, they cannot give a good distance estimate and resist the negative effects from the errors. In this paper we will propose new methods to improve its performance.

**HAUSDORFF DISTANCE**

Given two limited point sets $A$ and $B$, the Hausdorff distance (HD) is a non-linear operator. It defines the difference between the template and reference images by computing the maximum distance between the points in $B$ and $A$, and the maximum distance between the points in $A$ and $B$. So the target recognition and location task can be fulfilled by overlapping the template on the reference image and comparing the overlapped area using the HD.

**Partial Hausdorff Distance**

Hausdorff distance is defined as (Huttenlocher et al., 1993; Rucklidge, 1997)

$$H(A, B) = \max \left\{ h(A, B), h(B, A) \right\}$$

$$h(A, B) = \max_{a \in A} d_B(a) \cdot d_B(a) = \min_{a \in A} d(a, b)$$

where $h(A, B)$ is the inverse distance between $A$ and $B$; $h(B, A)$ is the forward distance; $d(a, b)$ is the Euclidean distance between two points $a$ and $b$. If $B$ is allowed to be transformed in some way, then the Hausdorff distance between the inversely transformed image $t^{-1}(A)$ and the template $B$ is defined as $H(t^{-1}(A), B)$.

Assuming there exists an acceptable match, the objective is to find such a transformation $\hat{t} \in T$ so that

$$H(t^{-1}(A), B) = \min_{\hat{t} \in T} H(t^{-1}(A), B)$$

In this paper, the transformation space $T$ is a concave subset of a six-dimensional Euclidean space $\mathbb{R}^6$ determined by six independent transformation parameters:

$$T = \{ t(d_x, d_y, s, \omega, \phi, \kappa) \mid -127 \leq d_x, d_y \leq 127, 8 \leq s \leq 1.2, 28 \leq \omega, \phi \leq .28 - \pi \leq \kappa \leq \pi \}$$

where $s$ is the relative scale factor of the reference image to the template; $d_x$ and $d_y$ are the sub-vectors of the base line zoomed out by the scale $s$; and $(\omega, \phi, \kappa)$ are the three rotation angles around $X$, $Y$, and $Z$ axes.

However, various distortions require that Hausdorff distance can conquer these negative effects. Huttenlocher (1993) suggested the partial Hausdorff distance $H_{KL}(A, B)$. It reads

$$H_{KL}(A, B) = \max\{h_k(A, B), h_l(B, A)\}$$

$$h_k(A, B) = K \cdot d_B(a)$$

where $K = f_k |A|, L = f_l |B|$ and $0 < f_k, f_l \leq 1$. The use of the fractions $f_k, f_l$ make $H_{KL}(A, B)$ robust to the disturbances due to occlusion, false and missing edges in some extent. But the results largely depend on choosing a good fraction value. Unless we have a priori information about the quantities of the noises, this metric cannot gain a good result. In addition, the PHD is not optimum for random noises because no an averaging operation is applied.

**Compute $d_B(a)$ by Distance Transformation**

The distance $d_B(a)$ is often the Euclidean distance. Precise computation of Euclidean distance is not necessary
because the edge map is affected by errors originated from sampling, noises, distortion and edge detection operators. This is fulfilled by distance transformation technique, and is equivalent to resample the Voronoi surfaces of the edge points in pixel resolution. The basic idea is that the global distance value is approximated by propagating the local distances between neighboring pixels. Therefore, a 2D edge map $B$ is converted to a distance map defined as

$$D_b(r,c) = \min_{b\in \Gamma} d((r,c),b)$$  

where $(r,c)$ are the row and column coordinates of a point in image. This equation assigns each non-edge pixel the distance to the nearest edge pixel and each edge pixel a zero value. We use the 5-7-11 distance transformation with a mask size of 5x5. The three local distance values are distance values between adjacent and diagonal pixels, respectively. The maximum difference between these approximate values and the Euclidean distance values is 2% (Borgerfors, 1986). The sequential DT has following steps:

- **Initialize** $D_b(r,c)$. Assign a zero to each edge pixel in the binary edge map, and a large positive to each non-edge pixel.
- **Divide** the square mask in Figure 1a into two smaller masks shown in Figure 1b, then scan the image two times: the forward mask scans from top-left to bottom right; and the backward mask scans from bottom right to top left.

Figure 1a shows a sample edge map, and the distance transformation result is illustrated in Figure 1b. A pixel is expressed using a smaller gray value if it has a larger distance value. Let $d_{b}(q)$ denote the value of a pixel $(r_{u},c_{u})$ in the distance map, that is $D_{b}[r_{u},c_{u}]$. The Hausdorff distance can be re-written as

$$H(A,B) = \max \{ \max_{a\in A} D_{b}[r_{u},c_{u}], \max_{b\in B} D_{b}[r_{u},c_{u}] \}. $$

**ADAPTIVE AND ROBUST HAUSDORFF DISTANCE**

Several modified Hausdorff distances, including M-HD, LTS-HD and $\alpha$-HD (Kwon, 1996), and MHD (Dubuisson, 1994), have been proposed to improve the robustness of the PHD. The three distance formulae proposed by Kwon (1996) are not sensitive to the parameters $f$, and thus can resist the disturbance from outliers and noises. So they are more robust and efficient than the PHD and MHD. However in theory, they estimated the mean of the whole sample instead of $d_{(\mu)} = \max_{i} d_i$, and thus is not an estimation of the original Hausdorff distance. In this section, we will propose two new formulae of the Hausdorff distance. The first one is call hybrid Hausdorff distance (HHD), which is a combination of the PHD and the MHD. The proof of the HHD of being a pseudo distance is given in Annex I. The second one is called robust reverse Hausdorff distance (RRHD), which gives an interval estimation of the $\alpha$. The RRHD is defined as

$$\hat{h}_s(A,B) = \frac{1}{s-r+1} \sum_{(i,s \leq i \leq s)} d_{(i)}$$

where $r = f_l \cdot |A|$, $s = f_h \cdot |A|$, $[f_l, f_h] \subseteq [0.5,1]$, and $f_h \in [f_l, f_h]$. The fractions $f_l, f_h$ are low and high fraction, respectively, such as $f_l = 0.8, f_h = 0.9$.

**Adaptive Robust Reverse Hausdorff Distance**

Compared with Kwon (1996)’s formulae, Eq. 7 is a more straightforward improvement to the PHD. However, the use of fixed fractions is short of the flexibility. We thus introduce an adaptive procedure to dynamically obtain a best interval estimation of $f_h$, and make the estimated distance values as close as possible to the actual distances.

**Definition of Distribution Function of the Order Statistic** $d_{(i)}$. The distance value $d_{b}(q)$ of any pixel $q (r_{q},c_{q})$ in the edge map can be obtained by probing the distance map $D_{b}$, that is, $d_{b}(q) = D_{b}[r_{q},c_{q}]$. If we
repeat such probing experiments \( n \) times, then a simple random sample is recorded, that is,\
\[ \{d_B(q_1), d_B(q_2), \ldots, d_B(q_n)\}, \]
where \( d_B \) denotes the corresponding random variable. Let the distribution be \( F_B \).

The inverse PHD \( h_B(A, B) \) actually calculates the \( f_R \) fraction of that distance sample. We know from mathematical statistics (Chen, 1981) that the \( k^{th} \) order statistic of the sample \( d_{(k)} \) has the following distribution:

\[
F_{d_{(k)}}(d) = \Pr(d_{(k)} < d) = \Pr(v_n(d) \geq k) = \sum_{i=k}^{n} C_n^i [F_B(d)]^i [1 - F_B(d)]^{n-i}
\]  

(8)

where, \( n = |A| \). The empirical frequency \( v_n(d) \) is the number of distances less than \( d \) among the \( n \) distance values, and has the binomial distribution \( B(n, F_B(d)) \). Since in different matching instances, the number of blunders is often different, and it is hard to determine the value of \( f_R \) so that all the large blunders are exactly rejected. We solve this problem by finding a best \( f_R \) in positive and negative directions along the distance curve so that the distance value approximates best.

**Distribution Function \( F_B \) Estimated from the Distance Map of the Template.** The probing points are assumed to be evenly distributed on the grid of \( B \). Therefore, \( n \) is equal to the number of grid points \( m \). The distribution function \( F_B(d) \) obeys by \( d_B \) can be computed using three steps:

1. Generation of the template distance map \( D_B \) of \( B \) using 5–7–11 distance transform.
2. Accumulation of the empirical frequency \( v_m(d) \). For the distance values \( 0, 1, 2, \ldots, d_{\text{max}} \) in \( D_B \), account empirical frequency \( v_m(0.2i), 0 \leq i \leq \lceil d_{\text{max}} / 0.2 \rceil \), where \( d_{\text{max}} \) is the maximum distance in \( D_B \).
3. Estimation of the sample distribution function \( F_{d_{(k)}}(d) \) . We can compute an approximate sample distribution function \( F_B(d) \) from \( F_m(d) = v_m(d) / m, 0 \leq d < +\infty \), and further distribution fit quality test can be done to obtain a continuous distribution function.

Figure 2 gives the sample distribution functions estimated from the distance map shown in Figure 1b. In addition, the estimate of the distribution function obeyed by \( d_B \) needs very complex computations. That is, we have to calculate a distance transform, empirical frequency and distribution function estimate for each search map \( A_t \) determined by each different transformation \( t \). That is, \( F_A \) changes with respect to \( t \) and is not discussed here.

**Inverse Partial Hausdorff Distance based on the Structures of Template Edges.** Two methods can be used to determine a best value of \( f_R \): (a) Inverse method. An initial value of \( f_R \) close to 1 (e.g., 0.9) is set, and then \( f_R \) is determined by gradually decreasing from the initial value. (b) Forward method. The fraction \( f_R \) is determined by gradually increasing from a small initial value \( f_R \) (e.g., 0.5). The steps used by the inverse method are as follows:

1. An interval estimate like \( [d_{(r)}, d_{(s)}] \) is calculated for the quartile \( d_f \) of \( f_R \) with \( t \leftarrow 0 \). For specified significance level \( \alpha \), we have to determine \( r, s \) so that the following condition is satisfied.

\[
P\{d_{(r)} > d_f\} = P\{d_{(s)} < d_f\} = \alpha / 2
\]  

(9)
If the distribution \( F_n \) is continuous at \( d_r \), then \( \hat{F}_n(d_r) = \hat{f} \). From Eq. 8, we have

\[
P[d_{(r)} > d_f] = P[v_n(d_f) < r] = \sum_{i=0}^{n-1} C_n f_i (1 - f)^{n-i} \tag{10a}
\]

\[
P[d_{(r)} < d_f] = P[v_n(d_f) \geq s] = \sum_{i=0}^{n} C_n f_i (1 - f)^{n-i} \tag{10b}
\]

Two methods can be used to obtain solutions to \( r \) and \( s \):

(a) Calculate distribution functions \( F_{d(j)} \) for the upper part of order statistics \( d_{(i)} \), \( k \geq n/2 \), then we have

\[
r = \arg \min_{k \in \{n/2 \cdots n\}} \left| P[d_{(i)} > d_f] - \sum_{i=0}^{k} C_n f_i (1 - f)^{n-i} \right| \tag{11a}
\]

\[
s = \arg \min_{k \in \{0 \cdots n-1\}} \left| P[d_{(i)} < d_f] - \sum_{i=0}^{k} C_n f_i (1 - f)^{n-i} \right| \tag{11b}
\]

(b) Because there are often thousands of edge points in \( A_n \), the binomial distribution can be approximated by normal distribution for a large sample (Chen, 1981), i.e.,

\[
\sum_{i=0}^{n-1} C_n f_i (1 - f)^{n-i} = \Phi \left( \frac{r - 1/2 - nf}{\sqrt{nf(1 - f)}} \right) \tag{12}
\]

To get the value of \( \alpha/2 \), we have

\[
r = nf + 1/2 + \mu_{\alpha/2} \sqrt{nf(1 - f)} \tag{13a}
\]

where \( \mu_{\alpha/2} \) is the \( \alpha \) / 2 quantile of the standard normal distribution. Similarly,

\[
s = nf - 1/2 - \mu_{\alpha/2} \sqrt{nf(1 - f)} \tag{13b}
\]

(2) Adaptively adjust the value of \( f \). Do the \( r \) test for a part of order statistics \( d_{(r)}, \ldots, d_{(s)} \) at the significance level of \( \beta (=0.05) \) assuming a linear regression hypothesis, and the slope \( u \) of the line is also calculated. Fisher’s approximate testing method is used since \( n \) is usually very large. If the linear hypothesis is accepted and the slope \( u \) is smaller than the overall slope \( u_{0i} = d_{[A]} / |A| \) of the \( d \sim n \) distance curve (discussed in the following section).

The part of order statistics within the \( 1 - \alpha \) confidence interval of the sample’s \( s/n \) fraction \( d_{iso} \) are rejected in testing, then \( f_s \leftarrow f^{(i)} \), and \( \hat{h}_{f_s} = \frac{1}{s - r + 1} \sum_{i \in [s-r+1]} d_{(i)} \) is the estimate to \( \hat{h}_{f_s} \), and stop the computation. Or \( t \leftarrow t + 1 \), \( f^{(i)} \leftarrow r/n \), if \( f^{(i)} > 0.5 \), then repeat Steps (1) and (2); or choose \( \hat{h}_{f_s} \leftarrow \hat{h}_f \) according to Eq. 7.

The forward method can use the same steps as the inverse method provided that the adjustment direction related with \( f \) is changed appropriately. We call \( \hat{h}_{f_s} \) the adaptive robust reverse Hausdorff distance (ARRHD). The above procedure is based on the following concrete observation: when the template is approximately mapped to the correct location in the reference image, the inliers distance values in the sample should change smoothly while the irregular changes and steep slopes will happen only at places where outliers occur. The experimental results have validated this idea as illustrated in Figure 3, which shows the \( d \sim n \) distance curve of the distance map given in Figure 1b. This \( d \sim n \) curve is obtained for a transformation of \( t = (98.766816, -101.82165, 894266, 0.76545, -0.081105, 0.79515) \), and illustrates the change of the order statistics \( d_{(i)}, k \in I_{01} \). The best fraction obtained is \( d_{0.552} \), and its confidence interval is \( [r, s] = [554, 590] \) at the significant level of 0.05. The ARRHD value reads \( \hat{h}_{0.552} = .744 \), and the overall slope of the curve is \( u_{0i} = .0267 \). Table 1 shows the adjustment procedure for the distance.

The adaptive distance \( \hat{h}_{f_s} \) is less sensitive to the fractions compared with the PHD since we calculated the empirical distribution function from the distance map of the template and thus fully utilize the distribution structures of the edge points. This scenario makes it possible to obtain the best fraction values dynamically and robustly, and thus produces a distance value closest to the true situation.

**Analysis of the \( d \sim n \) Distance Curve.** At Step (2) in the previous subsection, we can not assure an optimal result by adjust the fraction \( f \) in a simple way. According to the experiences with testing typical distance curves, this adaptive adjustment procedure should be discussed in eight cases (see Table 2) when considering if the linear hypothesis is satisfied around three fractions \( d_{iso}, d_1 \) and \( d_{iso} \) and if the slopes of the regressed lines are small enough.

Obviously, Case 2 is optimum. \( f \) should be increased in Case 1. While it should be decreased in Cases 7 and 8. In Case 3, we should increase \( f \) first, and decrease it if failed. Cases 4, 5 and 6 cannot be decided easily and have to
be processed according to the testing results. Usually, Cases 3 to 6 will occur when there are translation and scaling between the reference and the template images. In Figure 4, a typical $d \sim n$ curve is shown when using a biased transformation of $\hat{t} = (-.55, .65, 1.015, 0, 0, 0)$. It is observed that frequent jumps occur along the curve, and in this situation, the use of Eq. 7 can produce good results.

![Figure 3. The $d \sim n$ distance curve](image)

**Table 1. Dynamic adjusting procedure of fraction $f$ in negative direction**

<table>
<thead>
<tr>
<th>t</th>
<th>$f^{(t)}$</th>
<th>$r^{(t)}$</th>
<th>$s^{(t)}$</th>
<th>slope $u^{(t)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.9516</td>
<td>628</td>
<td>650</td>
<td>.1081</td>
</tr>
<tr>
<td>1</td>
<td>.9361</td>
<td>616</td>
<td>641</td>
<td>.1118</td>
</tr>
<tr>
<td>2</td>
<td>.9183</td>
<td>603</td>
<td>631</td>
<td>.0809</td>
</tr>
<tr>
<td>3</td>
<td>.898</td>
<td>588</td>
<td>619</td>
<td>.0606</td>
</tr>
<tr>
<td>4</td>
<td>.876</td>
<td>572</td>
<td>605</td>
<td>.0413</td>
</tr>
<tr>
<td>5</td>
<td>.852</td>
<td>554</td>
<td>590</td>
<td>.0125</td>
</tr>
</tbody>
</table>

**Table 2. Adaptive adjusting methods of fraction $f$**

<table>
<thead>
<tr>
<th>direction</th>
<th>$d_{r/n}$</th>
<th>$d_f$</th>
<th>$d_{s/n}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>2</td>
<td>P</td>
<td>P</td>
<td>F</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>P</td>
<td>F</td>
</tr>
<tr>
<td>5</td>
<td>P</td>
<td>F</td>
<td>P</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>F</td>
<td>P</td>
</tr>
<tr>
<td>7</td>
<td>P</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>

[Note]: P indicates the satisfaction of the linear hypothesis and the condition of a small slope; while F indicates the failure of the satisfaction. <= indicates $f \leftarrow r/n$; -> indicates $f \leftarrow s/n$; ? Indicates the uncertainty.

**Effectiveness of ARRHD.** To evaluate the effectiveness of the proposed ARRHD measure, two groups of aerial reference and template images shown in Figures 5 and 6 are used for the testing. We compared the theoretical distances and the ARRHD output distances when the values of the six transformation parameters are moderately biased from their known accurate values. Table 3 gives a partial of typical numerical results for the two reference images $I_1$ and $I_2$ at two locations respectively. In the second column of Table 3, the locations of the center points of
those templates on the reference images are given. By comparing the fourth and fifth columns, we can observe that the distance biases are less than 0.4 pixel. Taking into account of the quantizing error of 0.1 pixel due to the 5-7-11 distance transformation, we can conclude that the ARRHD distances are coincide with the theoretical values.

These experimental results show that within moderate biases (approximately 4 to 8 times the parameter incremental steps) of the parameters from their true values, the ARRHD can correctly measure the numerical relationship of the distance biases between the reference and template images. This proves that this new distance can be tolerant to the changes between the reference and template images due to different illumination, projection distortion and occlusion. And thus is an effective measure for image matching.

Table 3. Comparing theoretical and ARRHD distance under biased transformations

<table>
<thead>
<tr>
<th>Reference image</th>
<th>Template locations</th>
<th>Biased transformation (i(d_e,d_f,s,\omega,\phi,\kappa))</th>
<th>Theoretical distances</th>
<th>ARRHD outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I_1)</td>
<td>(M_a)</td>
<td>((0, 0, 1, 0, 0, 0))</td>
<td>.000000</td>
<td>.000000</td>
</tr>
<tr>
<td></td>
<td>(I_1): (0, 0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(I_2): (-109, 10)</td>
<td>((1, -1, -0.03, -0.02, -0.03))</td>
<td>2.695798</td>
<td>2.537965</td>
</tr>
<tr>
<td></td>
<td></td>
<td>((-1.5, -1.5, -0.03, -0.04, -0.03))</td>
<td>2.541987</td>
<td>2.392989</td>
</tr>
<tr>
<td>(M_b)</td>
<td>(I_1): (100, -100)</td>
<td>((100, -100, 1, 0, 0, 0))</td>
<td>.000000</td>
<td>0.047619</td>
</tr>
<tr>
<td></td>
<td>(I_2): (-9, -90)</td>
<td>((101.5, -100+1.5, 1.04, -0.03, -0.03 0.04))</td>
<td>4.039637</td>
<td>3.826056</td>
</tr>
<tr>
<td></td>
<td></td>
<td>((100-1, -100+1, 1.04, -0.03, -0.04, -0.03))</td>
<td>3.724525</td>
<td>3.564184</td>
</tr>
<tr>
<td>(I_2)</td>
<td>(M_c)</td>
<td>((0, 0, 1, 0, 0, 0))</td>
<td>.000000</td>
<td>.000000</td>
</tr>
<tr>
<td></td>
<td>(I_2): (0, 0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(I_1): (109, -10)</td>
<td>((1.5, 1.5, 1.04, -0.03, -0.04, -0.03))</td>
<td>2.793388</td>
<td>2.573560</td>
</tr>
<tr>
<td></td>
<td></td>
<td>((-1.5, -1.5, -1.03, -0.04, -0.04, -0.03))</td>
<td>2.541987</td>
<td>2.367827</td>
</tr>
<tr>
<td>(M_d)</td>
<td>(I_2): (-100, 100)</td>
<td>((-100+1.5, 100+1.5, 1.02, -0.03, -0.03, -0.04))</td>
<td>2.360432</td>
<td>2.293194</td>
</tr>
<tr>
<td></td>
<td>(I_1): (9, 90)</td>
<td>((-100+1.5, 100-1.5, -0.03, -0.03, -0.03, -0.05))</td>
<td>2.693595</td>
<td>2.504007</td>
</tr>
</tbody>
</table>

Adaptive and Robust Hausdorff Distance

Obviously, the APHD is still a type of PHD and thus has all the properties described in Hut (1993). So all the speeding up techniques can be used in the searching. As the inverse distance determines the similarity between some part of \(A\) and \(B\), the transformation that achieves the minimum value gives the best position of \(A\) relative to \(B\).

Therefore, a good substitute is to replace the directionless distance with \(\hat{h}_{\mu}\), that is called adaptive robust Hausdorff distance (ARHD) \(H(\tau^{-1}(A),B^-)\). But false match may occur when the blocks with little edges in the reference...
image are matched to the template. So the forward distance is vital in rejecting this transformation. We tackle this problem by designing appropriate thresholds.

**Determination of the Rejection Thresholds.** Thresholds are often used to discard incorrect transformations and their neighbors to relieve unnecessary computations. In addition, the hierarchical matching procedure using pyramids will propagate the errors in high levels to low levels, and the invalid transformations also lead to extra computations in the low levels. So it is viable to discard invalid transformations as soon as possible. Let \( \sigma_0 \) denote the specified positioning accuracy. The difference threshold \( \tau_1 \) of the edge point numbers and the inverse distance threshold \( \tau_R \) can be determined as follows.

1. The threshold \( \tau_1 \) of point number difference between the template and the search map. As mentioned before, there will be false matches at regions where the edge points are too dense or too loose in the reference image when using a metric that has only forward or inverse distance. This shows the number of edge point in both the template and the search map should be nearly equal for a correct transformation. Therefore, a threshold \( \tau_1 \) \((0 < \tau_1 < 1)\) may be defined so that a qualified transformation \( t \) should satisfy

\[
\frac{|A| - |B|}{|B|} \leq \tau_1
\]

(14)

The threshold \( \tau_1 \) can compensate the drawback of the partial distance metric to some extent without the need of more computational burden. Of course, we know from Rucklidge (1997) that the partial distance is well satisfactory for some special matching tasks.
Inverse distance threshold $\tau_R$. The best inverse fraction $f_R$ has been determined. We can approximate $F_{\hat{h}}$ using the distribution function of $d_{(nf_k)}$ since the order statistics are not independent and their joint distribution is very complex.

(a) Assuming the required location accuracy to be $\sigma_0(=2)$ and let $k = nf_R$, the distribution function obeyed by $\hat{h}/f_{\hat{h}}$ may be computed using Eq. 8, i.e.,

$$F_{\hat{h}}(d) = 1 - \Phi\left(\frac{k - 1/2 - nf_{\hat{h}}(d)}{\sqrt{nf_{\hat{h}}(d)(1 - F_{\hat{h}}(d))}}\right) = \gamma$$

(b) Determine $\tau_1$ so that $F_{\hat{h}}(\tau_1) = P[d_{(k)} < \tau_1] = \gamma$ is satisfied, that is $\tau_1 = F_{\hat{h}}^{-1}(\gamma)$, where $\gamma(<0.5)$ is a small probability.

Discarding the larger solution, we have

$$\tau_1 = F_{\hat{h}}^{-1}\left(\frac{2k + u_{\gamma}^2 - 1 + u_{\gamma}\sqrt{4k - 4k^2/n + 4k/n + u_{\gamma}^2 - 2 - 1/n}}{2(n + u_{\gamma}^2)}\right)$$

(c) Choose a value of $\tau_R$ within the interval of $[\sigma_0, \tau_1]$. It is easily observed that the estimation of $\tau_R$ also utilized the information about the distribution structure of the edge points in the template.

A Further Assumption. Assuming a correct transformation $\hat{t}$, we can know that the most edge points in $\hat{t}^{-1}(A_j)$ and $B$ have a one to one mapping within a very small distance. If $d_B(a)$ obeys the uniform distribution $U[0,\sigma_0]$, then the distribution function obeyed by $\hat{h}/f_{\hat{h}}$ can be approximated by

$$F_{\hat{h}}(d) = P[d_{(k)} < d] = P[v_{\sigma_0}(d) \geq k] = \sum_{i=k}^\infty C_i^k (1 - d/\sigma_0)^{i-k}$$

Thus given a $\tau_2$, we have $F_{\hat{h}}(\tau_2) = P[d_{(k)} < \tau_2] = \lambda$, that is, $\tau_2 = F_{\hat{h}}^{-1}(\lambda)$, where $\lambda(>0.5)$ is a large probability. Discarding the larger solution, we get

$$\tau_2 = \sigma_0 \cdot 2k + u_{\gamma}^2 - 1 - u_{\gamma}\sqrt{4k - 4k^2/n + 4k/n + u_{\gamma}^2 - 2 - 1/n}$$

If $\lambda = 0.95, n = 671, f_R = 0.852$, then we have $\tau_2 = \sigma_0 \cdot 0.825 = 1.65$. So the rejection threshold $\tau$ can be within the interval of $[\tau_2, \tau_1] = [1.65, 4.1]$. More testing results show that a wrong matching will be resulted when the distances are larger than 3, and thus we have $\tau \in [1, 3]$ in general.

MATCHING PROCEDURE AND TESTING RESULTS

The matching procedure using the adaptive Hausdorff distance has several steps as follows:
1. Image pre-processing by applying adaptive filtering (Saint-Marc and Chen, 1991) on the image and the template.
2. Edge detection. The binary features are obtained by detecting edges using Canny operator (Canny, 1986).
3. Distance transformation. The serial 5–7–11 distance transformation is applied on the edge map of the template to obtain a good distance map.
4. Estimate the distance distribution function of the template to get the empirical distribution function $F_B$ obeyed by the random variable $d_B$.
5. Template matching. Obtain the best estimate of the inverse distance by iteratively compute the inverse fraction \( f_R \) and do linear regression and statistical tests. Determine the best transformation \( \hat{t} \) by searching the transformation space \( \mathcal{T} \), i.e.,

\[
\hat{t} = \arg \min_{t \in \mathcal{T}} \hat{h}_t(A, B)
\]  

(20)

Such a solution of \( \hat{t} \) gives the best attitude parameters relating the template to the image, and also determines the best relative orientation of the template relative to other known targets in the image.

The two reference images \( I_1 \) (see Figure 5) and \( I_2 \) (see Figure 6) are composed of a stereo pair. It is should be noted that the origins of the coordinate systems of both the reference and template images are located at their image centers respectively. Among the templates of \( I_1 \), \( M_1 \sim M_3 \) are taken from \( I_1 \) itself, and \( M_4 \sim M_9 \) are from \( I_2 \). And among the templates of \( I_2 \), \( M_{10} \sim M_{15} \) are taken from \( I_1 \), and \( M_{16} \sim M_{18} \) are from itself. Table 4 gives a part of their known transformations and the results obtained by matching using the ARHD. Parts of testing results are also illustrated in Figures 7 and 8.

<table>
<thead>
<tr>
<th>No.</th>
<th>Known transformations</th>
<th>Transformations Found by ARHD</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_1 )</td>
<td>((-100, -100, 1, -1, 1, \pi/4))</td>
<td>((-100.009, -100.001, 1.0001, -100003, 100002, 7853))</td>
<td>Figure 7</td>
</tr>
<tr>
<td>( M_3 )</td>
<td>((-100, 100, 1, 1, -1, \pi/2))</td>
<td>((-99.993, 99.991, 1.1006, .0997, -.0998, 1.5708))</td>
<td>-</td>
</tr>
<tr>
<td>( M_5 )</td>
<td>((0, 0, 1, 0, 0, 0))</td>
<td>((-0.0008, -0.00512, .9999, -.000013, .000043, .000059))</td>
<td>-</td>
</tr>
<tr>
<td>( M_7 )</td>
<td>((100, -100, 9, -1, 1, \pi/4))</td>
<td>((99.773, -100.168, .8995, 100.029, -10013, 785))</td>
<td>-</td>
</tr>
<tr>
<td>( M_{12} )</td>
<td>((-100, 100, 1, -1, 1, 3\pi/4))</td>
<td>((-100.1449, 100.203, .992, 105. -100.007, 2.370))</td>
<td>Figure 8</td>
</tr>
<tr>
<td>( M_{14} )</td>
<td>((0, 0, 1, 1, -1, -1, 3\pi/4))</td>
<td>((.1787, -2109, 1.100008, -1098, 1002, -2.356))</td>
<td>-</td>
</tr>
<tr>
<td>( M_{16} )</td>
<td>((100, -100, 1, 1, -1, -1, \pi/4))</td>
<td>((99.979, -99.972, 10998, -.0992, -1005, -785))</td>
<td>-</td>
</tr>
<tr>
<td>( M_{18} )</td>
<td>((100, 100, 1, 1, -1, -1, \pi/4))</td>
<td>((100.005, 99.999, .9999, .009999, -10013, .7853))</td>
<td>-</td>
</tr>
</tbody>
</table>

**CONCLUSIONS**

The conventional partial Hausdorff distance uses fixed fraction to eliminate effects of noises and outliers. However, the ratios of outliers are often different for different matching instances. It is hard to determine an appropriate value of the fraction in advance so that all the outliers are exempt from the estimate perfectly. So the use of the partial Hausdorff distance has great limitations. We propose to compute an empirical distribution function obeyed by the distance variable from the distance map of each template. This obtains sufficient information about the distribution structure of edge points in the templates, and makes the adaptive Hausdorff distance insensitive to the fraction values. The fractions are determined robustly by analyzing the distance curves and adjusting search directions. This method can get the best distances closest to the actual situations by adaptively calculate best fraction values, which are robust. The experiments using aerial images show that ARHD has good performance in matching
templates in heavily blurred and complex backgrounds with pose change, scale change, geometric distortion and partial occlusion. Thus the ARHD is both effective, and can be a good index of locating accuracy.

\[ H(A, B) = H(A, B) + \frac{1}{|A|} \sum_{a \in A} d_B(a) + \frac{1}{|B|} \sum_{b \in B} d_A(b) \] (I.0)

**Proposition**: Let \( A \) and \( B \) denote two closed subsets that are non-empty and bounded in Euclidean space \( \mathbb{R}^2 \). The hybrid Hausdorff distance (HHD) defined in Equation I.0 is a pseudo metric.

\[ d(A, C) \leq d(A, B) + d(B, C) + \rho(B) \] (I.1)

where

\[ d(A, C) = \min_{a \in A, c \in C} d(a, c), \quad \rho(B) = \max_{b_1, b_2 \in B} d(b_1, b_2). \]

If \( A \) and \( B \) are sets each with a single point, i.e., \( A = \{a\}, B = \{b\} \), then Equation I.1 becomes

\[ d_c(a) \leq d(a, b) + d_c(b) \leq d(a, b) + H(B, C) \] (I.2)

If \( A \) is a set with a single point, i.e., \( A = \{a\} \), then Equation I.1 becomes

\[ d_c(a) \leq d_B(a) + d(B, C) + \rho(B) \Rightarrow \frac{1}{|A|} \sum_{a \in A} d_c(a) \leq \frac{1}{|A|} \sum_{a \in A} d_B(a) + d(B, C) + \rho(B) \] (I.3)
\[ d(B, C) = \min_{b \in B} d_c(b) \leq \frac{1}{|B|} \sum_{b \in B} d_c(b) \]  

Substituting Equation 1.4 into Equation 1.3, we have

\[ \frac{1}{|A|} \sum_{a \in A} d_c(a) \leq \frac{1}{|A|} \sum_{a \in A} d_b(a) + \frac{1}{|B|} \sum_{b \in B} d_c(b) + \rho(B) \]  

Similarly,

\[ \frac{1}{|C|} \sum_{c \in C} d_A(c) \leq \frac{1}{|C|} \sum_{c \in C} d_b(c) + \frac{1}{|B|} \sum_{b \in B} d_A(b) + \rho(B) \]

According to Equation 1.2, \( \forall a \in A, b \in B \),

\[ d_c(a) \leq \min_{b \in B} d(a, b) + H(B, C) = d_b(a) + H(B, C) \]

\[ \Rightarrow \max_{a \in A} d_c(a) \leq \max_{a \in A} d_b(a) + H(B, C) \leq H(A, B) + H(B, C) \]

We can assume

\[ H(A, C) = \max\{\max_{a \in A} d_c(a), \max_{c \in C} d_A(c)\} = \max_{a \in A} d_c(a) \]

with any arbitrary. Then

\[ H_1(A, C) = \frac{1}{|A|} \sum_{a \in A} d_c(a) + \frac{1}{|C|} \sum_{c \in C} d_A(c) + \max_{a \in A} d_c(a) \]

Substituted with Equations 1.5, 1.6 and 1.7, Equation 1.8 reads

\[ H_1(A, C) \leq \frac{1}{|A|} \sum_{a \in A} d_b(a) + \frac{1}{|B|} \sum_{b \in B} d_c(b) + \rho(B) \frac{1}{|C|} \sum_{c \in C} d_c(c) + \frac{1}{|B|} \sum_{b \in B} d_A(b) + \rho(B) + H(A, B) + H(B, C) \]

\[ = \frac{1}{|A|} \sum_{a \in A} d_b(a) + \frac{1}{|B|} \sum_{b \in B} d_c(b) + H(A, B) + \frac{1}{|B|} \sum_{b \in B} d_c(b) + \frac{1}{|C|} \sum_{c \in C} d_b(c) + H(B, C) + 2\rho(B) \]

\[ \Rightarrow H_1(A, C) \leq H_1(A, B) + H_1(B, C) + 2\rho(B) \]

(1) If \( \rho(B) = 0 \), i.e., \( B \) is always a single point set, then the triangular inequality is satisfied;

(2) If \( \rho(B) \neq 0 \), then the triangular inequality is not satisfied.

So the triangular inequality is not satisfied by the HHD.

**REFERENCES**


