Robust Hausdorff distance matching algorithms using pyramidal structures

Oh-Kyu Kwon, Dong-Gyu Sim, Rae-Hong Park*

Department of Electronic Engineering, Sogang University, C.P.O. Box 1142, Seoul 100-611, South Korea

Received 10 August 1999; received in revised form 4 May 2000; accepted 4 August 2000

Abstract

This paper proposes two Hausdorff distance (HD) matching algorithms, in which robust HD measures are implemented in pyramidal structures. By computer simulations, the matching performance of the conventional HD measures and the proposed robust HD matching algorithms using pyramidal structures is compared, with real images which are degraded by noise and occlusions.

Keywords: Hausdorff distance (HD); Object matching; M-estimation; Pyramidal structure; Least trimmed squares (LTS)

1. Introduction

Object matching in two-dimensional images has been an important topic in computer vision, object recognition, and image analysis [1–3]. The performance of the matching method depends on the type of the features used, matching measure criterion, and so on. Low-level matching algorithms, i.e., algorithms using a distance transform (DT) [4–6] and a Hausdorff distance (HD) [7–9] have been investigated because they are simple and insensitive to changes of image characteristics.

The robust statistics of regression analysis [10,11] was applied to computation of the HD measures for object matching, resulting in two robust HD measures: M-HD based on M-estimation and least trimmed squares–HD (LTS–HD) based on LTS [12]. These robust approaches yield the correct results, even though the input data are severely corrupted. However, they require high computational complexity in computing the minimum distance between two edge point sets. The DT was employed to compute the minimum distance from one point in the model image to the nearest point in the test image. Also these HD matching approaches require high computational complexity for searching for the best-fit matching position. To reduce the computational load, an efficient method based on multi-resolution tessellation [13] was proposed.

This paper proposes two robust HD matching algorithms using pyramidal structures. Under the translational motion, the accuracy of the proposed HD matching methods remain almost the same, with greatly reduced computational complexity, as that of the conventional HD matching methods in which each position is checked whether it is the optimal position or not.

The rest of the paper is structured as follows. In Section 2, conventional HD object matching algorithms are explained and in Section 3, the proposed robust HD object matching algorithms using pyramidal structures are presented. Experimental results for images containing noise and occlusions are shown in Section 4 and Section 5 concludes.

2. Convetional HD object matching algorithms

The HD computes distance values between two sets of edge points extracted from the object model and a test image.
The HD measure is sensitive to degradations such as noise and occlusions, so improved methods have been proposed for image or object matching. Huttenlocher et al. proposed a partial HD measure [7] for comparing partial portions of images containing severe occlusions or degradations. The directed distance of the partial HD is defined as

$$h_k(A, B) = K_{\text{ran}}^{th} d_p(a),$$

where $K_{\text{ran}}^{th}$ denotes the $K^{th}$ ranked value of $d_p(a)$, and the parameter $f$ is defined by the partial fraction such as $f = K/N_A$, with $N_A$ representing the number of points in the set $A$. This matching scheme is one of ranked-order statistics methods and yields good results in comparing images that contain outliers and occlusions. Dubuisson and Jain proposed the MHD based on the average distance value in comparing the synthetic images contaminated by four types of noise [9], in which the directed MHD $h_{\text{MHD}}(A, B)$ was defined as

$$h_{\text{MHD}}(A, B) = \frac{1}{N_A} \sum_{a \in A} d_p(a).$$

This matching scheme can optimally estimate a matching position in images with zero-mean Gaussian noise and does not require a parameter, whereas the partial HD scheme uses one user-specified parameter.

### 3. Proposed robust HD object matching algorithms using pyramidal structures

#### 3.1. HD measures based on robust statistics

Huttenlocher et al. proposed the partial HD measure in comparing partial portions of images containing severe occlusions or degradation [7]. This HD measure needs one parameter, partial fraction $f$, $0 < f < 1.0$. By modifying the HD based on the ranked-order statistics, Azencott et al. proposed the censored HD (CHD) measure in comparing binary images [8]. The CHD measure needs two parameters and requires the high computational load. The MHD measure [9] does not require a parameter, however, its matching performance is not good compared with that of the partial HD and the CHD, because it employs the summation operator over all distances, some of which might be computed from outliers.

To obtain more accurate object matching results, two HD measures based on the robust statistics such as $M$-estimation and the LTS were proposed [12]. The directed distance $h_M(A, B)$ of the M-HD based on $M$-estimation was realized by replacing the Euclidean distance by the cost function that can eliminate outliers [14]. $M$-estimation minimizes the summation of a cost function $\rho$. The directed distance $h_M(A, B)$ is defined as

$$h_M(A, B) = \frac{1}{N_A} \sum_{a \in A} \rho(d_p(a)), \tag{3}$$

where $d_p(a)$ denotes the smallest distance of point set $B$ at point $a$. The cost function $\rho$ is convex and symmetric, and has a unique minimum value at zero. In our experiments, we use the cost function $\rho$ defined by

$$\rho(x) = \begin{cases} |x|, & |x| \leq \tau, \\ \tau, & |x| > \tau, \end{cases} \tag{4}$$

where $\tau$ is a threshold to eliminate outliers, so the outliers yielding large errors are discarded. In M-HD, because the matching performance depends on the parameter $\tau$, it is important to determine $\tau$ appropriately.

Based on the LTS scheme [11,15], the directed distance $h_{\text{LTS}}(A, B)$ of the LTS-HD [12] is defined by a linear combination of order statistics

$$h_{\text{LTS}}(A, B) = \frac{1}{H} \sum_{i=1}^{n} d_p(a_i^0), \tag{5}$$

where $H$ denotes $h \times N_A$, as in the partial HD case, and $d_p(a_i^0)$ represents the $i$th distance value in the sorted sequence $d_p(x_1^0) \leq d_p(x_2^0) \leq \cdots \leq d_p(x_{N_A}^0)$. The measure $h_{\text{LTS}}(A, B)$ is minimized by remaining distance values after large distance values are eliminated. So, even if the object is occluded or degraded by noise or distortion, this matching algorithm yields good results.

#### 3.2. HHD algorithm based on pyramidal structures

Object matching algorithms using robust HD measures need high computational complexity. To reduce the computational complexity, we implement conventional HD matching schemes such as partial HD, MHD, M-HD, and LTS-HD using pyramidal structures.

Proposed HD matching algorithms are constructed using hierarchical structures [16], as shown in Fig. 1, in which the directed HDs between the DT map and the edge image are shown at each level, with the number of levels $L$ equal to three. Fig. 1 shows the construction of the three-level DT map pyramid and its edge pyramid in the hierarchical HD (HHD) matching algorithm, where $2^l$ denotes $2:1$ decimation in horizontal and vertical directions.

The distance value $D(i,j)$ at $(i,j)$ of a given binary image contains a number, corresponding to the distance value to the closest point in an image. Let the sizes of the object model and the test image be $p \times q$ and $P \times Q$, respectively, where $p \leq P$ and $q \leq Q$.

Under translational motions, the two directed HDs at level $l$ are computed by using the correlations of the edge image and the DT map:

$$h_l(A_t, B_t \oplus t_j) = \max_{(i,j)} \sum_{(i,j)} D_{f}^{l}(i - x_i^l, j - y_j^l), \tag{6}$$
The RHD measures based on M-estimation and LTS scheme can be proposed using the pyramidal structure as

\[ h_{\text{MT}}(A_t, B_t \oplus t_t) = \frac{1}{N_{A_t}} \sum_{(i,j) \in A_t} \rho_i \left( E_{i,j}^{\text{det}}(i, j) D_{i,j}^{\text{det}} \left( \frac{x_i}{2^l}, \frac{y_i}{2^l} - \frac{x_t}{2^l}, \frac{y_t}{2^l} \right) \right), \]  

(10)

\[ h_{\text{LTSI}}(A_t, B_t \oplus t_t) = \frac{1}{H_{I_{t}} \sum_{k=1}^{H_{I_{t}}} \left( E_{i,j}^{\text{det}}(i, j) D_{i,j}^{\text{det}} \left( \frac{x_i}{2^l}, \frac{y_i}{2^l}, \frac{x_t}{2^l}, \frac{y_t}{2^l} \right) \right)_k}, \]  

(11)

where the cost function \( \rho_i \) is convex and symmetric at level \( l \) and defined by

\[ \rho_i(x) = \begin{cases} \frac{|x|}{\xi_i}, & |x| \leq \xi_i, \\ \xi_i, & |x| > \xi_i \end{cases} \]

with \( \xi_i \) representing a threshold selected experimentally to eliminate outliers at level \( l \). \( H_{I_t} \) denotes \( h_t \times N_{A_t} \), where \( h_t \) signifies a partial fraction.

The proposed algorithm makes use of downsampling by a factor of 2, so the maximum error of positions detected between successive levels is one pixel in a finer level. Since rigid hierarchical structures used in pyramid can be sensitive to the position of the edges relative to the sampling grid, the threshold selection method [16] is adopted in the proposed matching algorithm. The proposed algorithm is based on a coarse-to-fine strategy, in which the position error could be corrected in the final level estimation, which prevents incorrect matching at low resolution.

The concept of thresholding can be explained as follows. The minimum HD value is found at level \( l \), and the

Fig. 1. Block diagram of the HD matching algorithm using pyramidal structures.

\[ h(B_t \oplus t_t, A_t) = \max_{(i,j)} E_{i,j}^{\text{det}} \left( i - \frac{x_i}{2^l}, j - \frac{y_i}{2^l} \right) D_{i,j}^{\text{det}}(i, j), \]  

(7)

\[ \frac{x_t}{2^l} \leq i_t < \frac{x_t + p}{2^l}, \quad \frac{y_t}{2^l} \leq j_t < \frac{y_t + q}{2^l}, \]  

where \( E_{i,j}^{\text{det}}(i, j) \) and \( D_{i,j}^{\text{det}}(i, j) \) denote the binary edge (DT map) values at level \( l \) of the test image and the object model, respectively. The translation \((x_t, y_t)\) at level \( l \) is assumed to be in the range from 0 to \((P - p) / 2^l - 1\), and from 0 to \((Q - q) / 2^l - 1\) in the \( x \)- and \( y \)-direction, respectively.

The directed HD values for the partial HD and MHD are modified by replacing the max operator in Eq. (6) by the rank operator \( K^\text{th} \) and the summation operator \( \Sigma \), respectively, then the directed forms of the partial HD and MHD at level \( l \) become

\[ h_{\text{Kd}}(A_t, B_t \oplus t_t) = K^{\text{th}}_{i,j} E_{i,j}^{\text{det}}(i, j) D_{i,j}^{\text{det}}(i, j), \]  

(8)

\[ h_{\text{MHD}}(A_t, B_t \oplus t_t) = \frac{1}{N_{A_t}} \sum_{(i,j) \in A_t} E_{i,j}^{\text{det}}(i, j) D_{i,j}^{\text{det}}(i, j), \]  

(9)

where \( N_{A_t} \) denotes the number of edge points in the set \( A_t \) at level \( l \). Because the MHD measure gives an optimal solution to Gaussian noise cases and is not robust to impulse noise, which will be observed in experimental results with noisy images, the HHHD matching algorithm using Eq. (9) as the directed HD measure does not yield good matching results. Thus, we use Eq. (8) as the directed HD measure.
positions at which the difference between the HD value and the minimum HD value is less than the threshold value $\tau_i$ are selected as the candidate positions, i.e., the positions that satisfy the following inequality are chosen as the candidate positions:

$$0 \leq H_l(A_l, B_l) - H_l^{\text{min}}(A_l, B_l) \leq \tau_i,$$

where $H_l(A_l, B_l)$ represents the HD value computed at level $l$ and $H_l^{\text{min}}(A_l, B_l)$ signifies the minimum HD value at level $l$, with the superscript $\text{min}$ denoting the minimum HD value at level $l$. This procedure is repeated recursively from the coarsest level ($l = L - 1$) down to the original level ($l = 0$).

Selection of the threshold value $\tau_i$ is an important problem in the proposed robust HD matching methods using pyramidal structures. Because the computation time depends on the number of candidate positions, which is related to the threshold value $\tau_i$, large $\tau_i$ requires high computational load in finding the optimal matching position. On the contrary, if $\tau$ is small, it takes less time, but the detected position may be incorrect.

4. Experimental results and discussions

To show the matching performance of the proposed robust HD matching methods using pyramidal structures, the performance of the conventional and proposed HD algorithms is compared, in terms of the accuracy of the matching position and the computation time, for several real images with various levels of noise, distortion, and occlusions. The matching position is defined by the position, detected by each HD measure, of the object model with respect to the test image. The conventional HD matching schemes require high computational complexity. The partial HD, MHD, $M$-HD, and LTS-HD are simulated as the conventional HD measures in performance comparison. In experiments, the $M$-HD and LTS-HD measures give a more accurate position than the partial HD, because of the average operation embedded in the robust HD measures, and produce better results than the MHD by effectively removing outliers caused by occlusions.

Figs. 2(a) and (b) show the binary $72 \times 48$ object model and the test Taxi image, respectively. The object model is extracted from the binary image obtained by the Canny edge operator. The origin (0,0) of the model Taxi image corresponds to position (70,81) of the test Taxi image.

To generate noisy binary images used in experiments, we randomly add uniform noise to the original noise-free image by flipping pixel values [9]. The amount of noise added is expressed in terms of the percentage $U$, defined by $U = 100R/(P \times Q)$, where $R$ denotes the number of pixels whose values are reversed, and $U$ represents the percentage of $R$ to the total number of pixels in the image, with $P$ and $Q$ signifying the number of rows and columns of a test image, respectively.

Fig. 2(c) shows the superimposed matching result of Fig. 2(a) and a noisy binary image Fig. 2(b) contaminated by uniform noise of $U = 12\%$. In Fig. 2(c), the proposed partial HD matching algorithm using pyramidal structures is used, with the parameter $f$ set to 0.8. Note that the parameter $f = 0.8$ gives the nearly correct matching
position of (68,79). Pyramidal realization may detect a position a little bit different from that detected by the original HD measure. The number of pyramid levels and the candidate threshold \( \tau \) are experimentally set to three and 2, respectively.

Table 1 shows results of the conventional and proposed HD matching algorithms for the real Taxi image. The matching positions estimated by the two robust HD measures (M-HD with \( \tau = 3, 4, 5 \), and LTS-HD with \( h = 0.6, 0.7, 0.8 \)) are correct, thus the superimposed matching results are identical. Also the proposed HD object matching methods with the same parameter values yield the correct results.

Fig. 3(a) shows the binary 72 \( \times \) 72 Road image and Fig. 3(b) shows the 256 \( \times \) 256 Road image contaminated by Gaussian noise (standard deviation \( \sigma = 30 \)) [8]. Also about 35\% pixels of the target portion of an input image are deleted. The identical matching result obtained by the two robust HD matching methods (M-HD and LTS-HD) using pyramidal structures is shown in Fig. 3(c), where the object model is superimposed on the test image.

Fig. 4 shows the RMS matching position error for the noisy Road image by the conventional and proposed HD matching methods, as a function of the corresponding parameter of each HD measure, with two different uniform noise levels (\( U = 12 \) and 20\%). As the uniform noise level \( U \) increases, the RMS position error increases.

In Fig. 4(a), where \( \tau \) is equal to 0, because there is no data which can measure the similarity between objects, the matching result is bad for all M-HD measures. Also for large \( \tau \), because the outliers are not effectively removed by the cost function \( \rho \), the RMS position error is large. Fig. 4(b) shows that if the parameter \( h \) of the LTS-HD measure is from 0.50 to 0.90, the RMS position error is 0, i.e., the matching results are correct. If \( h \) is larger than 0.90, the LTS-HD measure yields.
incorrect results, because the outliers are not effectively removed.

Figs. 5(a) and (b) show the $82 \times 112$ model image and the $256 \times 256$ test image, respectively. The object models are extracted from the binary images obtained by the Canny edge operator. The origin (0,0) of the Aerial images corresponds to position (62,70) of the corresponding test images.

Fig. 5(c) shows the superimposed matching result of Fig. 5(a) and a noisy binary image Fig. 5(b) contaminated by uniform noise of $U = 20\%$. In Fig. 5(c), the proposed HHD matching algorithm is used, in which the LTS-HD is employed as the HD measure, with the parameter $h$ set to 0.75. The number of pyramid levels is experimentally set to three ($L = 3$).

Fig. 6 show the scatter diagrams of the matching positions of the Aerial image detected by the conventional method and the proposed HHD matching algorithm with two noise levels of $U = 6\%$ and $20\%$, respectively. The total number of experiments is equal to 200. With $U = 6\%$, the number of positions correctly detected by the conventional method and the proposed HHD matching algorithm are equal to 39 and 194, respectively. It is noted that the efficiency of the proposed method for noisy images is greatly improved compared with the conventional method.
Fig. 6. Scatter diagrams of matching positions detected by the conventional method and LTS-HHD with two noise levels (U = 6 and 20%) (aerial image).

Table 2
Comparison of the computational complexity of conventional and proposed HHD matching algorithms (L-level pyramidal structures)

<table>
<thead>
<tr>
<th></th>
<th>Conventional HD matching</th>
<th>HHD matching algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computational complexity</td>
<td>((N_A + N_B)[(P - p)(Q - q)])</td>
<td>((N_A + N_B)\left[\sum_{i=0}^{L-2} \frac{1}{2^{i}} N_i + \frac{1}{4^{L-1}}(P - p)(Q - q)\right])</td>
</tr>
<tr>
<td>Additional calculation</td>
<td>For generation of a DT map pyramid and an edge pyramid</td>
<td></td>
</tr>
</tbody>
</table>

The parameters of the \(M\)-HD and LTS-HD methods cannot be accurately determined. The parameters of the robust HD measures depend on the amount of noise and various kinds of objects/noise. Appropriate \(\tau\) and \(h\) of \(M\)-HD and LTS-HD are experimentally set to about 4–5 and 0.7–0.8, respectively. The number of pyramid levels \(L\) and the threshold \(\tau_l\) at level \(l\) are experimentally set to three and 2\(^{l}\), respectively.

With the similar matching performance, the number of comparisons by the proposed matching algorithms for calculation of the minimum distance is reduced by a factor of about 40, compared with that of the conventional methods, neglecting the computation required for construction of the DT map pyramid and the edge map pyramid.

Table 2 shows the comparison of the computational complexity of the conventional method and the proposed HHD matching algorithm. Computational complexity of the conventional method and the proposed one under translational motions is compared, with the \(P \times q\) \((P \times Q)\) object model (test image) containing \(N_A(N_B)\) edge points. In the conventional HD matching scheme, the number of comparisons for calculation of the minimum distance is equal to \((N_A + N_B)[(P - p) \times (Q - q)]\). In the proposed HHD method, at level \(l\) the number of comparisons for calculation of the minimum distance is equal to \((1/2^{l-1})[N_A + N_B(1/4^{L-1})P - p)(Q - q)\). Also, the proposed HHD matching algorithm needs extra computation for construction of the DT map and edge pyramids.
In experiments with the $256 \times 190$ Taxi image in Fig. 2(c), on MIPS R4400 (175 MHz) workstation, it takes about 122 s with the $M$-HD method, whereas about three seconds with the proposed $M$-HD matching algorithm using pyramidal structures, with the same performance, where the three-level pyramid is used. In experiments with the $256 \times 256$ Road image in Fig. 3(c), it takes about 350 s by the LTS-HD method, whereas about 9 s by the proposed LTS–HD matching algorithm using pyramidal structures, where the three-level pyramid is used. In experiments with the $256 \times 256$ Aerial image in Fig. 5(c), it takes about 370 s by the LTS-HD method, whereas about 9 s by the proposed LTS–HD matching algorithm using pyramidal structures, where the three-level pyramid is used.

5. Conclusions

This paper proposes two robust HD matching methods using pyramidal structures to greatly reduce the computational complexity. The computation time of the proposed HD matching methods is reduced by a factor of about 40, with the comparable performance. The effectiveness of the proposed HD matching methods is tested with various images and noise environments. Further research will focus on the applications of the HD algorithms to various nonideal cases.

Acknowledgements

Authors are gratefully acknowledging the financial support by Agency for Defence Development and by Automatic Control Research Center (ACRC), Seoul National University.

References

Laboratory of the Center for Automation Research, University of Maryland, College Park, Maryland, as a visiting associate professor. He received a scholarship from the Korean Government, Ministry of Education, from 1979 to 1983, and a Post-Doctoral Fellowship from the Korea Science and Engineering Foundation (KOSEF) in 1990. He received the Academic Award in 1987 from the Korea Institute of Telematics and Electronics (KITE). Also he received the First Sogang Academic Award in 1997 and the Professor Achievement Excellence Award in 1998 and 1999, all from Sogang University. He served as Editor for the KITE Journal of Electronics Engineering in 1995–1996. His current research interests are computer vision, pattern recognition, and video communication.