Hausdorff Distance with k-Nearest Neighbors*

Jun Wang and Ying Tan**

Key Laboratory of Machine Perception (Ministry of Education), Peking University
Department of Machine Intelligence, School of EECS, Peking University, P.R. China
wangjun@cis.pku.edu.cn, ytan@pku.edu.cn

Abstract. Hausdorff distance (HD) is an useful measurement to determine the extent to which one shape is similar to another, which is one of the most important problems in pattern recognition, computer vision and image analysis. Howeverm, HD is sensitive to outliers. Many researchers proposed modifications of HD. HD and its modifications are all based on computing the distance from each point in the model image to its nearest point in the test image, collectively called nearest neighbor based Hausdorff distances (NNHDs). In this paper, we propose modifications of Hausdorff distance measurements by using k-nearest neighbors (kNN). We use the average distance from each point in the model image to its kNN in the test image to replace the NN procedures of NNHDs and obtain the Hausdorff distance based on kNN, named kNNHDs. When $k = 1$, kNNHDs are equal to NNHDs. kNNHDs inherit the properties of outliers tolerance from the prototypes in NNHDs and are more tolerant to noise.

Keywords: Hausdorff distance, k-Nearest neighbor, face recognition.

1 Hausdorff Distance

To determine the extent to which one shape is similar to another is one of the most important problems in pattern recognition, computer vision and image analysis. The performance of the matching method can be viewed as the measurements to determine the difference between shapes, which is the similarity between two objects based on their shape attributes [12,8]. The criterion of determining these measurements includes the efficiency of computation, robust and reasonable results, and firmly theoretical basis [12]. Hausdorff distance (HD) is such a measurement that has been widely used in pattern recognition [10,15,16].

HD is a non-linear operator to measure the mismatch of two given sets [6,8]. Given two finite point sets $M = m_1, m_2, ..., m_m$ (representing a model image) and $T = t_1, t_2, ..., t_t$ (representing a test image), Hausdorff distance is defined as [8]:

$$H(M,T) = \max\left(h(M,T), h(T,M)\right), \tag{1}$$

and where

$$h(M,T) = \max_{a \in M} \min_{b \in T} \parallel a - b \parallel. \tag{2}$$

* This work was supported by the National Natural Science Foundation of China (NSFC), under grant number 61170057 and 60875080.

** Corresponding author.

Y. Tan, Y. Shi, and Z. Ji (Eds.): ICSI 2012, Part II, LNCS 7332, pp. 272–281, 2012.

The function H(M,T) is called *undirected Hausdorff distance* and the function h(M,T) is called *directed Hausdorff distance*.

Hausdorff distance measures the similarity extent of a point set to another. Not like most of other shape comparison methods, Hausdorff distance is not based on finding the corresponding mode. Thus, it is more tolerant in the perturbations of points locations, since it considers proximity more than exact superposition. However, Hausdorff distance is very sensitive to outliers [8]. In certain circumstances, given two sets of points A and B, B is set A plus one single point which is far from other points in A. Thus, the Hausdorff distance of A and B is determined by the distance of the single point to other parts. The sensitive to outliers is fatal to measurement of similarity. Therefore, many researchers propose modifications of the original Hausdorff distance to reduce the influence of outliers. Several typical modifications are enumerated as below:

1. RHD
   Huttenlocher et al. propose the Ranked Hausdorff distance (RHD) [8]. They propose a definition of "p^{th} distance" for the model points ($1 \leq p \leq |M|$) which takes the p^{th} ranked point of M. This definition automatically selects the p^{th} "best matching" points of M to ignore obvious outliers and minimizes the *directed Hausdorff distance*. They define the *directed Hausdorff distance* of RHD as:
   $$h^p_{RHD}(M,T) = p^{th} \max_{a \in M} \min_{b \in T} \parallel a - b \parallel, \tag{3}$$
   which compute the p^{th} maximum distance. When p=1, RHD equals to HD.
2. MHD
   Dubuisson et al. propose the modified Hausdorff distance (MHD) [4]. Unlike previous item, they use the average distance from each point in M to its nearest point in T to instead one single point. The *directed Hausdorff distance* is defined as:
   $$h_{MHD}(M,T) = \frac{1}{|M|} \sum_{a \in M} \min_{b \in T} \parallel a - b \parallel \tag{4}$$
   where $|M|$ is the amount of points in M.
3. M-HD and LTS-HD
   Kwon et al. propose two modifications of Hausdorff distance [9,13] based on the robust statistics such as M-estimation and the LTS scheme.
   The *directed Hausdorff distance* of M-HD is defined based on M-estimation to replace the Euclidean distance by a cost function ρ to eliminate outliers.
   $$h_{M\_HD}(M,T) = \frac{1}{|M|} \sum_{a \in M} \rho(\min_{b \in T} \parallel a - b \parallel). \tag{5}$$
   The cost function ρ is convex and symmetric, and has a unique minimum value at zero. Thus, the large error outliers are discarded. The cost function ρ is defined as:
   $$\rho(x) = \begin{cases} |x|, & |x| \leq \tau, \\ \tau, & |x| > \tau, \end{cases} \tag{6}$$
   where τ is a threshold to eliminate the outliers.

Based on the LTS scheme, the *directed Hausdorff distance* of LTS-HD is defined by a linear combination of order statistics.

$$h_{LTS}(M,T) = \frac{1}{H}\sum_{i=1}^{H} d_T(a)_i \qquad (7)$$

where H denotes the amount of points participate statistics $H \leq |M|$, and $d_T(a)_i$ denotes the i^{th} minimal distance value in the sorted sequence.

These measurements treat noise as "true" feature points and have no idea to avoid the influence of noise. In practice, the feature image obtained from original image contains many noise. If there exists noises in images with some probability, then the measurements would be inaccurate with some probability. Therefore, it is important to reduce the noise influence in Hausdorff distance computation. In previous work, only Zhao et al.'s work involves this aspect [17].

2 Hausdorff Distance with k-Nearest Neighbors

In this paper, We propose modifications of Hausdorff distance measurements based on k-nearest neighbors. At first, we notice that in the computational procedure, HD, RHD, MHD, M-HD and LTS-HD all contain same procedure

$$\min_{a,b\in T} \parallel a-b \parallel (a \in M),$$

which computes the distance from each point in M to its nearest neighbor in T. This is the nearest neighbor (NN) problem and can be optimized by computing the distance transform [8]. For purposes of discussion, these five measurements are collectively called Hausdorff distances based on nearest neighbor (NNHDs). The nearest neighbor (NN) problem aims to find the closest point of a given test point among a set of points. The k-nearest neighbor (kNN) problem can be considered as an extension of NN, which aims to find k closest points of a given test point among a set of points. The k-nearest neighbor approach is an important technique for pattern recognition [5]. The influence of noise — the points appear on T but do not belong to target object, is improperly treated as the nearest neighbor of the point in M, which will mislead the measurement. Therefore, we use the information of k-nearest points in T for each point in M to replace the single nearest point. The average of the distances to the k-nearest points is more robust than the distance to one single point.

2.1 Principle

As usual, the proposed measurements take two finite point sets M as a model image and T as a test image. The image is an n by m array of which each element has a value 0 or 1, referred as background or feature (foreground) point, respectively. The feature points of T are denoted by F_T. Let $f_d(a,T)$ denote the d^{th} closet feature point in F_T of $a \in F_M$ ($\|\cdot\|$ denotes the Euclidean metric).

$$f_d(a,T) = \mathop{\arg\min}_{b\in F_T - \bigcup_{i=1}^{d-1} f_i(a,T)} \|a-b\|^2, a \in F_M. \qquad (8)$$

Let $NN_k(a,T)$ denote the k-nearest points in T of point a.

$$NN_k(a,T) = \bigcup_{d=1}^{k} f_d(a,T), a \in F_M. \tag{9}$$

There are many effective methods to compute the k-nearest neighbor problem [3,11]. Hence, it is easy to compute the k-nearest neighbor in T of each point in M, which will be discussed in more detail in following section. The average distance of point a to its k-nearest neighbor feature points in T is defined as:

$$Average_k(a,T) = \frac{1}{k} \sum_{b \in NN_k(a,T)} \|a - b\| \tag{10}$$

We use the average distance of k-nearest points to replace the distance to the nearest point in the computation of Hausdorff distance. Thus, we propose modifications of Hausdorff distance measurements based on k-nearest neighbors (kNNHDs), as follows:

1. k-HD is defined as

$$H_{HD}^{k}(M,T) = \max\left(h_{HD}^{k}(M,T), h_{HD}^{k}(T,M)\right), \tag{11}$$

and

$$h_{HD}^{k}(M,T) = \max_{a \in M}\left(Average_k(a,T)\right), \tag{12}$$

where k denotes k-nearest neighbors.

2. k-RHD is defined as

$$H_{RHD}^{p,k}(M,T) = \max\left(h_{RHD}^{p,k}(M,T), h_{RHD}^{p,k}(T,M)\right), \tag{13}$$

and

$$h_{RHD}^{p,k}(M,T) = p^{th} \max_{a \in M} Average_k(a,T), \tag{14}$$

where k denotes k-nearest neighbors and p denotes the p^{th} maximum distance.

3. k-MHD is defined as

$$H_{MHD}^{k}(M,T) = \max\left((h_{MHD}^{k}(M,T), h_{MHD}^{k}(T,M)\right), \tag{15}$$

and

$$h_{MHD}^{k}(M,T) = \frac{1}{|M|} \sum_{m \in M} Average_k(a,T), \tag{16}$$

where k denotes k-nearest neighbors.

4. k-M-HD is defined as

$$H_{M-HD}^{k}(M,T) = \max\left(h_{M\_HD}^{k}(M,T), h_{M\_HD}^{k}(T,M)\right), \tag{17}$$

and

$$h_{M\_HD}^{k}(M,T) = \frac{1}{|M|} \sum_{m \in M} M\_Average_k(a,T), \tag{18}$$

where k denotes k-nearest neighbors.

$M\_Average_k(a,T)$ denotes the average distance to kNN of a restricted by the cost function ρ. There are two forms of $M\_Average_k(a,T)$ definition:

$$M\_Average_k(a,T) = \rho(Average_k(a,T)) \tag{19}$$

or

$$M'\_Average_k(a,T) = \frac{1}{k} \sum_{b \in NN_k(a,T)} \rho(\|a-b\|) \tag{20}$$

The former one more conform the original intention of M-estimation. We choose (19) for the *directed Hausdorff distance* of k-M-HD.

The cost function ρ is defined as:

$$\rho(x) = \begin{cases} |x|, & |x| \le \tau, \\ \tau, & |x| > \tau, \end{cases} \tag{21}$$

where τ is a threshold to eliminate the outliers.

5. k-LTS-HD is defined as

$$H^k_{LTS}(M,T) = \max\left(h^k_{LTS}(M,T), h^k_{LTS}(T,M)\right), \tag{22}$$

and

$$h^k_{LTS}(M,T) = \frac{1}{H} \sum_{i=1}^{H} d^{k,i}_{LTS}(M,T), \tag{23}$$

where k denotes k-nearest neighbors.

$d^{k,i}_{LTS}(M,T)$ denotes the i^{th} minimal distance value in the sorted sequence, and H denotes the amount of points participate statistics $H \le |M|$.

$$d^{k,i}_{LTS}(M,T) = i^{th} \min\left(\bigcup_{a \in M} Average_k(a,T)\right). \tag{24}$$

In all these measurements, the inner parts of the *directed Hausdorff distance* are the average distance to its k-nearest neighbors. These measurements are the kNN version of NNHDs, collectively called kNNHDs (Hausdorff distance based on k-nearest neighbors). It should be noticed, when k = 1, the k-nearest neighbor is equal to nearest neighbor, and kNNHDs are equal to NNHDs, respectively.

3 Implementation

The efficiency of the measurement computation is also an important factor. Here, we implement kNNHDs in a efficient way. Huttenlocher et al. use the distance transform to optimize the computation of Hausdorff distance [8].

Algorithm 1 Compute $h^k_{HD}(M, T)$

Require: model image M, test image T and order k

compute the kNNT of T, and obtain k arrays $A_1 ... A_k$;
for i = 2 to k **do**
    $A_1 = A_1 + A_i$; //matrix operation
end for
$A_1 = A_1 \cdot M$; //matrix operation
$value$ = the maximum element of A_1;
$value = value/k$;
return $value$;

Similarly, we compute the k-nearest feature points for each point of the test image. We propose a k-nearest neighbor transform (kNNT) for binary images in [3]. The goal of kNNT is to assign each point in the binary image with its k-nearest feature points. The computational complexity of constructing kNNT for a binary image is $O(Nk^2)$ (N is the size of the image) with using $O(Nk)$ space [3]. Due to limitations on space, we only give the pseudo-code algorithm of the *directed Hausdorff distance* of k-HD. The implementations of other algorithms of kNNHDs are similar to this one.

Totally, $O(Nk^2 + Nk + N + N) = O(Nk^2)$ (M and T have same image size N). The other computational complexities of NNHDs and kNNHDs can also be analyzed, as follows:

Table 1. Computational complexity of measurments

Measurements	HD	RHD	MHD	M-HD	LTS-HD
NNHDs	$O(N)$	$O(Np)$	$O(N)$	$O(N)$	$O(NH)$
kNNHDs	$O(Nk^2)$	$O(Nk^2)$	$O(Nk^2)$	$O(Nk^2 + NH)$	$O(Nk^2 + Np)$

As Table.1 shows, the computational complexities of kNNHDs are similar to those of NNHDs, which are all linear with the image size. The only difference is that the computational complexities of kNNHDs are based on $O(Nk^2)$, which comes from the KNNT computation. Therefore, although the computational complexities of kNNHDs are a bit higher than those of NNHDs, there are still at same level.

4 Experiments

4.1 Synthetic Images

We used the binary images of the English letters as the synthetic images. First, we generated images of ten letters from "A" to "J"; next, and we use random noise to destroy these images with probability from 1% to 10% (Fig.1). Zhao et al. use similar English letters as synthetic images in [17] and they only use one kind of lower noise images. The experiment strategy is that: we compute the distance from the original images to the images with variant noise probability by the proposed measurements; for each level noise images, the right recognized letters is count as good and ten groups of recognition rates for each measurements are obtained.

Fig.2 shows that kNNHDs all present better performance than NNHDs. In Fig.2(a), HD and k-HD are all sensitive to outliers and show poor performance. At most time,

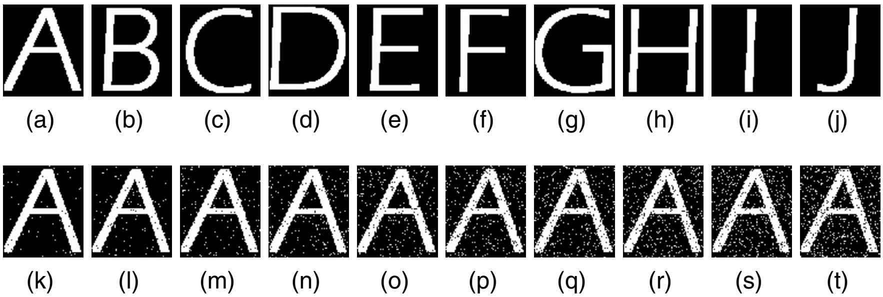

Fig. 1. Synthetic images of English letters. (a) — (j) the original image of corresponding letters; (k) — (t) demonstration of letter "A" destroyed by random noise with probability from 1% to 10%.

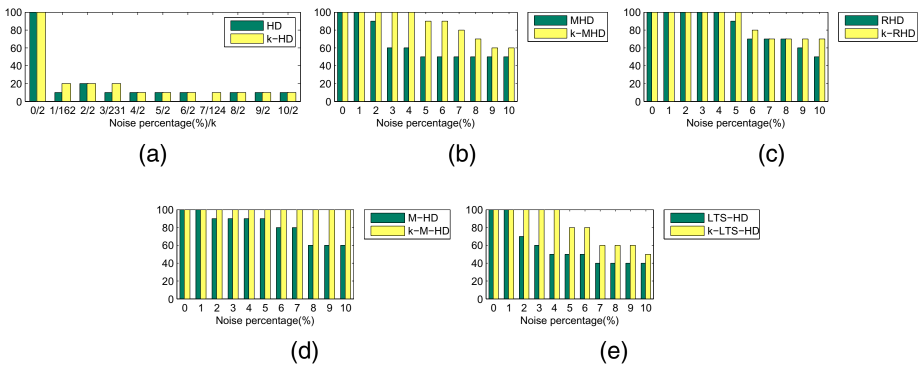

Fig. 2. Results on the synthetic images. (a) Comparison between HD and k-HD, variant k for different level noise images is shown in the figure ; (b) Comparison between MHD and k-MHD, k = 40; (c) Comparison between RHD and k-RHD, k = 155, p = 1/4 amount of total feature points; (d) Comparison between M-HD and k-M-HD, k=6, τ=4.5; (e) Comparison between LTS-HD and k-LTS-HD, k= 110, H = 90% of amount of feature points.

the results of k-HD are same as HD, and k-HD only has a slight improvement than HD at several levels of noise images with different k. The other four of kNNHDs achieve the highest recognitions rates at a single "k". Fig.2(b)(c)(d)(e) show the NN version and kNN version of MHD, RHD, M-HD and LTS-HD all effectively reduce the affection of outliers as we mentioned before. In Fig.2(b), by increasing noise, the recognition rate of MHD drops quickly, and the recognition rate of k-MHD is 60% with 10% of noise. In Fig.2(c), RHD represents better than MHD and shows poor performance after the noise upper than 5%. At the same time, the recognition of k-RHD is 70% with 10% of noise. In Fig.2(d), k-M-HD shows the best performance than all others — its recognition rate are 100% at all level noise percentage. In Fig.2(e), k-LTS-HD shows significant improvement compared with LTS-HD.

4.2 Face Recognition

We examined kNNHDs on real objects recognition. Many researchers use Hausdorff distance in face recognition [6,7,10,14,15,16]. We compared kNNHDs with NNHDs by face recognition on two favorite databases, the ORL database of faces [1] and the Yale faces database [2]. Images of the two databases are processed by same procedures to obtain the edge feature images(Fig.3, Fig.4):

1. Locate the face part of each original image and cut the redundant part, then normalize image to same size;
2. Use the Sobel filter on face image to get the edge image [6,15];
3. Threshold the image with dynamic threshold value, ensure the amount of feature points approximates 20% of the total edge image.

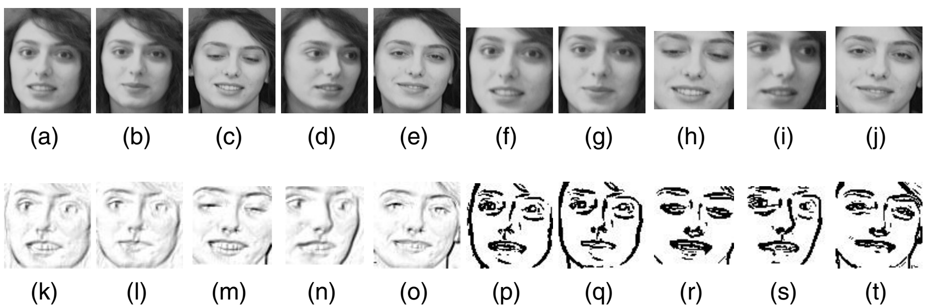

Fig. 3. Sample images of the ORL database of faces. (a) — (e) Original images of same individiaul; (f) — (j) Region of interest of face images; (k) — (o) Edge image extracted by Sobel filter; (p) — (t) Edge feature after threshold.

Next, we use kNNHDs and NNHDs to measure the similarity of the edge images of faces in proper order. The classify method is Nearest Neighbor [5], and the experimental strategy is Leave-One-Out. As shown in Fig.5 and Fig.6, the recognition rates of kNNHDs are better than those of NNHDs on both databases, expect k-MHD keeps same as MHD on the ORL database of faces (Fig.5(b)). The results of HD are poor compared with others and k-HD has a small increment than HD (Fig.5(a), Fig.6(a)). The results of k-RHD and k-M-HD show significant improvements than RHD and M-HD on both databases (Fig.5(c)(d), Fig.6(c)(d)). k-LTS-HD is better on the Yale faces database than on the ORL database of faces.

These results illustrate that kNNHDs promote the recognition rates in various degree. On the whole, kNNHDs, eliminating more noises from feature extraction, are more capable to measure the similarity of objects. Actually, in the research on the face recognition, many other information is also considered to improve the recognition rates, such as face gradient [16], spacial eigen-weighted [10] and shape contexture [15], etc.

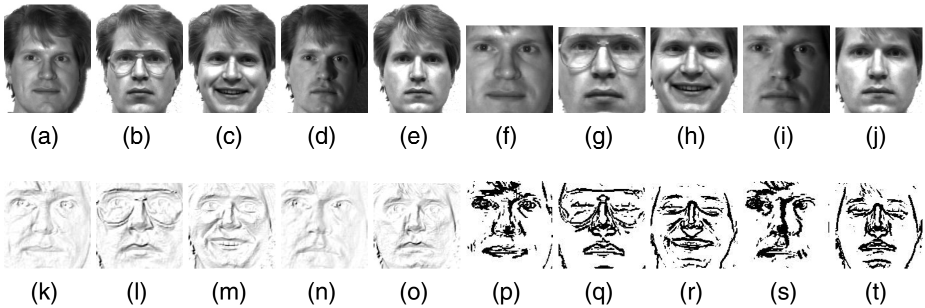

Fig. 4. Sample images of the Yale faces database. (a) — (e) Original images of same individiaul; (f) — (j) Region of interest of face images; (k) — (o) Edge image extracted by Sobel filter; (p) — (t) Edge feature after threshold.

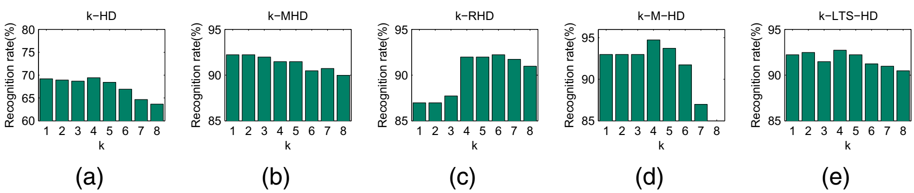

Fig. 5. The recognition rates of kNNHDs on the ORL database of faces with variable k. When k = 1, the results are of NNHDs.

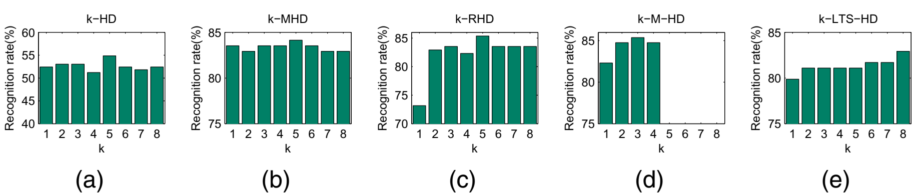

Fig. 6. The recognition rates of kNNHDs on the Yale faces database with variable k. When k = 1, the results are of NNHDs.

5 Conclusion

In this paper, we propose modifications of Hausdorff distance measurements by using k-nearest neighbors (kNN). Experiments on synthetic images and face recognition show that kNNHDs performs better than NNHDs, respectively. kNNHDs have several characters:(1) kNNHDs inherit the properties of reducing the sensitivity to outliers from the prototypes in NNHDs; (2) kNNHDs are more tolerable to noise; (3) kNNHDs are easy to compute. Therefore, kNNHDs are favorable measurements for pattern recognition and image matching. The methods based on NNHDs can be easily modified to adopt kNNHDs, which can benefit substantially from this modification. Although, we illustrate kNNHDs with Euclidean metric here. Other metrics, such as the city-block (L_1), the chessboard (L_∞), the octagonal, can all be employed by kNNHDs for different situations.

References

1. The ORL database of faces. The Oliver Research Laboratry in Cambridge, UK, http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
2. The Yale database of faces, Yale University, http://cvc.yale.edu/projects/yalefaces/yalefaces.html
3. Authors. k-nearest neighbor transform for binary images. In: CVPR 2012, Submission ID 129, Supplied as additional material knndt.pdf (2012)
4. Dubuisson, M.-P., Jain, A.: A modified hausdorff distance for object matching. In: Proceedings of the 12th IAPR International Conference on Pattern Recognition, vol. 1, pp. 566–568 (October 1994)
5. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley Interscience (2000)
6. Gao, Y., Leung, M.: Face recognition using line edge map. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(6), 764–779 (2002)
7. Gao, Y., Leung, M.: Line segment hausdorff distance on face matching. Pattern Recognition 35(2), 361–371 (2002)
8. Huttenlocher, D., Klanderman, G., Rucklidge, W.: Comparing images using the hausdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence 15(9), 850–863 (1993)
9. Kwon, O.-K., Sim, D.-G., Park, R.-H.: Robust hausdorff distance matching algorithms using pyramidal structures. Pattern Recognition 34(10), 2005–2013 (2001)
10. Lin, K.-H., Lam, K.-M., Siu, W.-C.: Spatially eigen-weighted hausdorff distances for human face recognition. Pattern Recognition 36(8), 1827–1834 (2003)
11. Nutanong, S., Jacox, E.H., Samet, H.: An incremental hausdorff distance calculation algorithm. Proc. VLDB Endow. 4, 506–517 (2011)
12. Rucklidge, W.: Efficient Visual Recognition Using the Hausdorff Distance. Springer-Verlag New York, Inc., Secaucus (1996)
13. Sim, D.-G., Kwon, O.-K., Park, R.-H.: Object matching algorithms using robust hausdorff distance measures. IEEE Transactions on Image Processing 8(3), 425–429 (1999)
14. Takács, B.: Comparing face images using the modified hausdorff distance. Pattern Recognition 31(12), 1873–1881 (1998)
15. Xie, X., Lam, K.-M.: Elastic shape-texture matching for human face recognition. Pattern Recogn. 41, 396–405 (2008)
16. Yang, C., Lai, S., Chang, L.: Hybrid image matching combining hausdorff distance with normalized gradient matching. Pattern Recognition 40(4), 1173–1181 (2007)
17. Zhao, C., Shi, W., Deng, Y.: A new hausdorff distance for image matching. Pattern Recogn. Lett. 26, 581–586 (2005)