Olszewska, Joanna Isabelle and Wilson, D.

Hausdorff-Distance Enhanced Matching of Scale Invariant Feature Transform Descriptors in Context of Image Querying

Original Citation

Olszewska, Joanna Isabelle and Wilson, D. (2012) Hausdorff-Distance Enhanced Matching of Scale Invariant Feature Transform Descriptors in Context of Image Querying. In: Proceedings of the IEEE 16th International Conference on Intelligent Engineering Systems. IEEE, Lisbon, Portugal, pp. 91-96. ISBN 978-1-4673-2694-0

Hausdorff-Distance Enhanced Matching of Scale Invariant Feature Transform Descriptors in Context of Image Querying

J.I. Olszewska and D. Wilson

School of Computing and Engineering, University of Huddersfield
Queensgate, Huddersfield, HD1 3DH, United Kingdom
{j.olszewska}@hud.ac.uk

Abstract— **Reliable and effective matching of visual descriptors is a key step for many vision applications, e.g. image retrieval. In this paper, we propose to integrate the Hausdorff distance matching together with our pairing algorithm, in order to obtain a robust while computationally efficient process of matching feature descriptors for image-to-image querying in standards datasets. For this purpose, Scale Invariant Feature Transform (SIFT) descriptors have been matched using our presented algorithm, followed by the computation of our related similarity measure. This approach has shown excellent performance in both retrieval accuracy and speed.**

I. Introduction

Image retrieval has attracted many attention during the last years [1] and is of the highest usefulness for public applications on the Web such as Google Image, etc. Two main approaches exist for image retrieval. One firstly requires image annotations or tags and then, the retrieval technique relies on the related semantic content of the images. The other technique is uniquely based on the image content, and consists of the direct search of pictures as pictures, without any semantic description nor keyword. In this paper, we adopt the second approach which is also called content-based image retrieval (CBIR). Indeed, CBIR constitutes an important aspect of image mining and retrieval because sometimes it is difficult to attach a word to picture. In particular, we focus on the "image-to-image querying". In that case, an image example which is external to the database feeds the retrieval system and results in a query for the most similar images [2]. In fact, an image in the database could not only differ from the image query by its orientation, position, or size, as it has been studied in the past [3], [4], but could also be another exemplar of the same category like in [5].

In opposite to the object recognition task [6], CBIR is characterized by:

- a broad domain, i.e. the database containing the images among which to search, with a very large number of classes;
- the absence of an explicit training for feature selection and classification;
- the difficulty to describe precisely the images because of the variability of sensing conditions and object states.

In fact, CBIR does not specifically involve image segmentation [7] or scene understanding [8] but requires the identification and detection of features, since the comparison between the query image and the data image is usually performed in the feature space.

The most popular features are the local descriptors because they are more robust to clutter, occlusion, and viewpoint changes than the main global features such as color histograms [9], [10], or color moments [11]. Some of the common local feature descriptors are the scale invariant feature transform (SIFT) [12], gradient location and orientation histogram (GLOH) [13], histogram of oriented gradient (HOG) [14], speeded-up robust features (SURF) [15], and DAISY [16]. These local distribution-based descriptors are computed in surrounding regions of extracted keypoints by means of interest point detectors such as Hessian detector [17], Harris [18], Laplacian of Gaussian (LoG) [19], Difference of Gaussian (DoG) [12], etc. or by dense sampling such as in [20].

In this work, we have merged three standard datasets to obtain a broad database for the image query system and we have selected scale invariant feature transform (SIFT) [12] which is the most useful local feature descriptor [1] to encode the image content and which is invariant to scale, rotation and translation.

For the purpose of the image-to-image querying, once the features are extracted from the query and the data images, respectively, the two obtained sets are then put in correspondence. This latter action is called matching [21] and could be based on the Hungarian algorithm [22], affine registration [23], voting algorithm [24], or various similarity measures [25], [26]. To decrease the computational burden, other approaches have been developed, e.g. the computation of distances such as the Euclidean one. However, Euclidean distance does not incorporate image spatial information and thus, is not enough discriminant in case of image deformation or noise. To overcome this shortcoming, improvements have been made like in [27] or other distances have been proposed

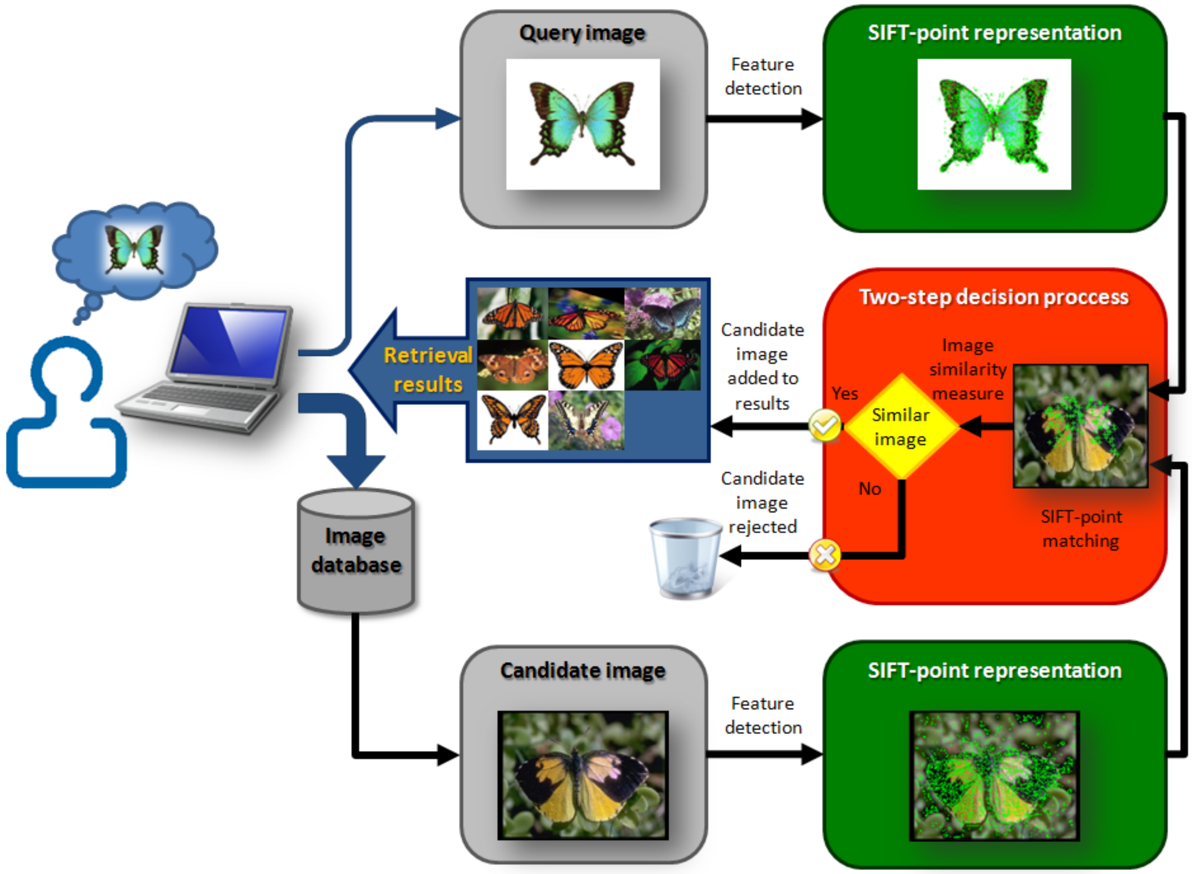

Fig. 1. Overview of our proposed architecture for image-to-image querying.

such as the Hausdorff distance [28], or its variations [29], [30]. Despite Hausdorff distance efficiency to cope with rotation, this distance is very sensitive to outliers and performs poorly with scale changes [31].

Hence, we propose in this paper an enhanced matching method more efficient in terms of both speed and accuracy for the content-based image retrieval than the separate computation of the traditional techniques.

The contributions of this paper are as follows:

- the matching technique which enhances our greedy pairing process with the Hausdorff distance combined with the city-block distance and the integration of a related image similarity measure into this Hausdorff-distance enhanced matching, leading to a very accurate, fast retrieval system;
- the architecture of the image-to-image querying process based on the proposed two-step SIFT-feature matching method.

The paper is structured as follows. In Section II, we present our image-to-image querying approach which is based on the matching of Scale Invariant Feature Transform (SIFT) descriptors using our Hausdorff - City-Block combined distance enhancing the pairing algorithm and on the related image similarity measure. The resulting image retrieval system has been successfully tested on a broad database containing real-world standard image datasets as reported and discussed in Section III. Conclusions are drawn up in Section IV.

II. Our Proposed Image-to-Image Querying System

In this section, we describe our proposed system (Fig. 1), which basically relies on the comparison of a query image with any image of a given database. The main steps are the detection of image features (Section II-A) and their matching (Section II-B). Indeed, the comparison between the query image and the candidate one is performed in the feature space in order to be more computationally effective and to be more robust towards noise and affine transformation, and is followed by the computation of an associated similarity measure (Section II-C).

A. Scale Invariant Feature Transform (SIFT) Descriptor Computation

To characterize an image, we adopt the Scale Invariant Feature Transform (SIFT) descriptors [12]. Indeed, SIFT constitutes a distinctive and fast local feature descriptor which is robust to rotation, translation, scale changes as well as some viewpoint variations.

The computation of SIFT consists of four stages:

1) The first stage is the *scale-space extrema detection* to search for scale-invariant features across all scales and image locations. These keypoints are identified by finding the maxima and minima of the Difference-of-Gaussian (DoG) function $D(x, y, \sigma)$ which provides a close approximation of the scale-normalized Laplacian of Gaussian (LoG). The DoG function can be computed from the difference of two nearby scales separated by a constant factor k, and is defined as follows

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y), \quad (1)$$

where $G(x, y, k\sigma)$ is the variable-scale bidimensional Gaussian and $I(x, y)$ is the input image.

2) In this step of *accurate keypoint localization*, keypoints identified in the previous step are selected based on their stability. Their location is calculated using a Taylor expansion up to the quadratic term of the scale-space function $D(x, y, \sigma)$, shifted to set the origin at the central sample point

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x}, \quad (2)$$

where D and its derivative are evaluated at the sample point and $\mathbf{x} = (x, y, \sigma)^T$ is the offset from this point. The location of the extremum $\hat{\mathbf{x}}$ is determined by taking its derivative with respect to $\mathbf{x}$ and setting it to zero, giving

$$\hat{\mathbf{x}} = -\frac{\partial^2 D}{\partial \mathbf{x}^2}^{-1} \frac{\partial D}{\partial \mathbf{x}}. \quad (3)$$

To eliminate extrema with low contrast, the function value of the extremum $D(\hat{\mathbf{x}})$ is calculated by substituting (3) in (2). This gives

$$D(\hat{\mathbf{x}}) = D + \frac{1}{2}\frac{\partial D^T}{\partial \mathbf{x}}\hat{\mathbf{x}}, \quad (4)$$

and then, all the keypoints whose extremum is below a threshold are rejected.
To remove keypoints which are poorly localized along edges, the principal curvatures of each keypoint are computed. All the keypoints whose ratio between the principal curvatures is above a threshold are then discarded.

3) The *orientation assignment* to each keypoint provides rotation invariance. First, gradient magnitude $m(x, y)$ and orientation $\theta(x, y)$ are calculated for each Gaussian smoothed image $L(x, y) = G(x, y, \sigma) * I(x, y)$, as

$$\begin{aligned} m = \quad & ((L(x+1, y) - L(x-1, y))^2 \\ & +(L(x, y+1) - L(x, y-1))^2)^{1/2}, \end{aligned} \quad (5)$$

$$\theta = tan^{-1}\left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}\right). \quad (6)$$

Then, the gradient orientations of sample points within a region around the keypoint form an orientation histogram with 36 bins which cover the 360 degrees. Each sample added to the histogram is weighted by its gradient magnitude and by a Gaussian-weighted circular window with a $\sigma_S = 1.5k$. The dominant peak in the histogram determines the orientation of the keypoint, whereas other local peaks within 80% of the highest peak are used to create others keypoints at the same scale and location but with these orientations.

4) Hence, a *keypoint descriptor* has been generated by assigning a scale, a location and an orientation to each stable keypoint. Finally, this feature vector is normalized to the unit length to decrease the effects of illumination changes, while its complexity is reduced in order to be less sensitive to shape distortion.

B. Hausdorff-Distance Enhanced Matching (HDEM)

Let us give the two finite sets of SIFT local descriptors

$$A \quad = \quad \{a_i \mid i \in [1; \#A]\}, \quad (7)$$
$$B \quad = \quad \{b_j \mid j \in [1; \#B]\}, \quad (8)$$

with $\#A = card(A)$, the number of keypoints characterizing the query image and $\#B = card(B)$, the number of keypoints characterizing the data image.

First, the Hausdorff distance $d_H(A, B)$ is computed as follows

$$d_H(A, B) = max\Big(d_h(A, B), d_h(B, A)\Big) \quad (9)$$

where $d_h(A, B)$ is the directed Hausdorff distance from A to B defined as

$$d_h(A, B) = \max_{a \in A} \min_{b \in B} d_P(a, b), \quad (10)$$

with $d_P(a, b)$, the Minkowski-form distance [32], based on the L_P norm, and defined as

$$d_P(a, b) = \left(\sum_k (a_k - b_k)^P\right)^{1/P}. \quad (11)$$

Then, the SIFT keypoints are matched using the Hausdorff distance calculated by (9). Hence, the directed Hausdorff distance $d_h(A, B)$ ranks each SIFT point of A based on its $d_P(a, b)$ distance to the nearest SIFT point of B and then uses the largest $d_P(a, b)$ distance which in fact results from the most mismatched point of A. The Hausdorff distance $d_H(A, B)$ is the maximum of the directed Hausdorff distances $d_h(A, B)$ and $d_h(B, A)$ [28]. Thus, every SIFT point of A is within the Hausdorff distance $d_H(A, B)$ of some SIFT point of B and vice versa. As there is no explicit pairing of points of A with points of B at this step, it could happen that many points of A may be matched to the same point of B.

Next, all the matched SIFT points are paired according to Algorithm 1 [33] involving the computed Hausdorff distance $d_H(A, B)$ which is used this time as the pairing threshold, avoiding thus an arbitrary threshold. The resulting doubly matched features constitute then the set M.

Algorithm 1 Matching Algorithm

Given $A' = A, B' = B, M = \varnothing$,
for all $a_i \in A'$ **do**
  for all $b_j \in B'$ **do**
    repeat
      if

$$\begin{aligned} & d_P(a_i, b_j) && = \min_{b \in B'} d_P(a_i, b) \\ \wedge\ & d_P(b_j, a_i) && = \min_{a \in A'} d_P(b_j, a) \\ \wedge\ & d_P(a_i, b_j) && \leq d_H(A, B) \\ \wedge\ & d_P(b_j, a_i) && \leq d_H(A, B) \end{aligned}$$

      then

$$\begin{aligned} & (a_i, b_j) \subset M \\ \wedge\ & A' = A \backslash \{a_i\} \wedge B' = B \backslash \{b_j\} \end{aligned}$$

      end if
    until $A' \neq \varnothing \vee B' \neq \varnothing$
  end for
end for
return M

C. Computing the Retrieval Results

The image similarity measure $d_S(A, B)$ is computed as follows

$$d_S(A, B) = \frac{\#M}{\frac{\#A + \#B}{2}}, \tag{12}$$

with A and B, the sets of SIFT features of the query and the candidate images, respectively, and M, the set of the double-matched ones [33].

The decision that a candidate image contains similar content to the query image one is taken based on the fact that the similarity measure $d_S(A, B)$ is above a given threshold. In the case when the similarity measure $d_S(A, B)$ is below the given threshold, the candidate image is rejected.

Our matching method's robustness to scaling, rotation, and translation and some changes in illumination is due the use of SIFT descriptors, while the one towards clutter, geometrical distortions and changes in appearance has been demonstrated in Section III.

III. Results Evaluation and Discussion

In order to test our matching approach for the image querying application, we have used three standards datasets, namely, the CalTech dataset - 101 categories [34], the Oxford Flower dataset [35] and the Caltech-UCSD Birds 200 database [36] that we have merged to obtain a broad domain of images suitable for public applications involving image retrieval such as search into Web.

In particular, the Caltech-101 database contains 9197 images of different objects such as butterflies, bikes, etc. and consists of 31 to 800 images per class. Image resolution is around 300x300 pixels.

The Oxford Flower database groups together about 8238 images of 102 types of common British flowers with a typical resolution of 667x500 pixels. Each category contains from 40 to 358 images. This dataset presents additional challenges of scale, pose and light variations as well as important class similarity.

The Caltech-ICSD Birds is an image dataset with 6033 photos of 200 bird species mostly from North America, which was primarily developed for subordinate categorization and which owns 20 to 39 images per class. The average resolution of these images is 375x500 pixels.

In this way, the images of our resulting database have different size and resolution as well as large inter-class similarities and intra-class variations. Hence, the difficulty of the image retrieval in this overall database is very high.

The carried out experiments use the query image presented in Fig. 2 as an example to find similar images in our overall database. This type of querying by image example is called approximate and it is the most common type of research in the Web [2]. In fact, the retrieved images should contain at least one object from the same category that the query one, whatever their poses and without being necessary exactly the same object.

All the experiments have been performed on a commercial computer with a processor Intel(R) Atom (TM) CPU N270 1.60 GHz, 1 Gb RAM and using MatLab (Mathworks, Inc.) software.

Some examples of the retrieved results using our method are illustrated in Fig. 3 (a) and some example of rejected images are presented in Fig. 3 (b).

To assess the accuracy of our retrieval system, we adopt the standard criterion as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{13}$$

with TP, true positive, TN, true negative, FP, false positive, and FN, false negative.

TABLE I
Image Retrieval Accuracy

Similarity [25]	Hausdorff [28]	HDEM
79%	82%	95%

We can observe in Table I that our system based on the embedded double matching is very accurate compared to the state-of-the art methods [25], [28]. Furthermore, our matching method outperforms even works such as [5] which uses multiple kernel classifiers to distinguish between object classes.

In order to measure the computational time performance of our matching approach in the context of the above-described image retrieval process, we carried out the following experiment. On one hand, we compute the Minkowski-form distance defined by (11) using the norm $L_P = L_1$. This leads to the City-Block distance or Manhattan distance $d_1(a, b)$. On the other hand, we compute the Minkowski-form distance using the norm $L_P = L_2$. Then, (11) defines the Euclidean distance $d_2(a, b)$.

In this experiment, we can observe that the computational time is c. 0.5 seconds and that the computational speed is twice faster when the L_1 norm is used in (11) rather than the L_2 norm. Thus, the calculation of the City Block distance to compute the enhanced Hausdorff distance in our matching technique allows faster image retrieval.

IV. Conclusions

In this work, the developed system for automatic image-to-image querying is based on the enhanced matching of Scale Invariant Feature Transform (SIFT) descriptors. Hence, the proposed matching relies on the Hausdorff distance enhancing a pairing process of the SIFT points detected in the query and the data images, respectively. A related similarity measure is then computed based on these doubly matched SIFT descriptors in order to retrieve images with similar content to a given example. As validated, our technique is very robust for image retrieval in large standard database while being computationally efficient.

References

[1] Z. Zhang and R. Zhang, *Multimedia Data Mining*, Chapman and Hall/CRC, 2009.

[2] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, December 2000.

[3] A.K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognition*, vol. 29, no. 8, pp. 1233–1244, August 1996.

[4] G. Mori, S. Belongie, and J. Malik, "Shape contexts enable efficient retrieval of similar shapes," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, December 2001, pp. I.723–I.730.

[5] M.-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in *Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing*, December 2008, pp. 722–729.

[6] R. Wood and J. I. Olszewska, "Lighting-variable AdaBoost based-on system for robust face detection," in *Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*, February 2012, pp. 494–497.

[7] J. I. Olszewska, *Unified Framework for Multi-Feature Active Contours*, Ph.D. thesis, UCL, 2009.

[8] J. I. Olszewska and T. L. McCluskey, "Ontology-coupled active contours for dynamic video scene understanding," in *Proceedings of the IEEE International Conference on Intelligent Engineering Systems*, June 2011, pp. 369–374.

[9] M.J. Swain and D.H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, November 1991.

[10] J. R. Smith and S.-F. Chang, "Single color extraction and image query," in *Proceedings of the IEEE International Conference on Image Processing*, October 1995, pp. III.528–III.531.

[11] M. Stricker and M. Orengo, "Similarity of color images," in *Proceedings of the SPIE International Conference on Storage and Retrieval for Image and Video Databases*, February 1995, vol. 2420, pp. 381–392.

[12] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, November 2004.

[13] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, October 2005.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, June 2005, pp. II.886–II.893.

[15] H. Bay, A. Ess, T. Tuytelaars, and L. van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, June 2008.

[16] E. Tola, V. Lepetit, and P. Fua, "DAISY: An efficient dense descriptor applied to wide baseline stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 815–830, May 2010.

[17] P. Beaudet, "Rationally invariant image operators," in *Proceedings of the Joint Conference on Pattern Recognition*, November 1978, pp. 579–583.

[18] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the ALVEY Vision Conference*, September 1988, pp. 147–151.

[19] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79–116, October 1998.

[20] H. Wang, P. Teng, and W. Liang, "Packed dense interest points for scene image retrieval," in *Proceedings of the IEEE International Conference on Image and Graphics*, August 2011, pp. 789–794.

[21] J. I. Olszewska, T. Mathes, C. De Vleeschouwer, J. Piater, and B. Macq, "Non-rigid object tracker based on a robust combination of parametric active contour and point distribution model," in *Proceedings of the SPIE Visual Communications and Image Processing Conference*, January 2007, pp. 6508–2A.

[22] D. Chitra, T. Manigandan, and N. Devarajan, "Shape matching and object recognition using dissimilarity measures with Hungarian algorithm," in *Proceedings of the International Conference on Man-Machine Systems*, October 2009, pp. IB1–IB6.

[23] J. Ho, M.-H. Yang, A. Rangarajan, and B. Vemuri, "A new affine registration algorithm for matching 2D point sets," in *Proceedings of the IEEE International Workshop on Applications of Computer Vision*, February 2007, p. 25.

[24] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 530–535, May 1997.

[25] C. Wolf, J.-M. Jolion, W. Kropatsch, and H. Bischof, "Content based image retrieval using interest points and texture features," in *Proceedings of the International Conference on Pattern Recognition*, September 2000, vol. 4, pp. 234–237.

[26] D. Zhang and G. Lu, "Evaluation of similarity measurement for image retrieval," in *Proceedings of the IEEE International Conference on Neural Networks and Signal Processing*, December 2003, pp. 928–931.

[27] J. Li and B.-L. Lu, "An adaptive image Euclidean distance," *Pattern Recognition*, vol. 42, no. 3, pp. 349–357, March 2009.

[28] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, September 1993.

[29] S.H. Kim and R.-H. Park, "An efficient algorithm for video sequence matching using the modified Hausdorff distance and the directed divergence," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 7, pp. 592–596, July 2002.

[30] C. Zhao, W. Shi, and Y. Deng, "A new Hausdorff distance for image matching," *Pattern Recognition Letters*, vol. 26, no. 5, pp. 581–586, April 2005.

[31] B. Gunsel and A. M. Tekalp, "Shape similarity matching for query-by-example," *Pattern Recognition*, vol. 31, no. 7, pp. 931–944, July 1998.

[32] S. Santini and R. Jain, "Similarity measures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 871–883, September 1999.

[33] T. Alqaisi, D. Gledhill, and J. I. Olszewska, "Embedded Double Matching of Local Descriptors for a Fast Automatic Recognition of Real-World Objects," in *Proceedings of the IEEE International Conference on Image Processing*, 2012.

[34] L. Fei-Fei, M. Andreetto, and M.A. Ranzato, "The Caltech - 101 Object Categories dataset, 2003.," Available online at: `http://www.vision.caltech.edu/feifeili/Datasets.htm`.

[35] M.-E. Nilsback and A. Zisserman, "Flower Dataset - 102 Category dataset, Oxford, 2006.," Available online at: `http://www.robots.ox.ac.uk/ vgg/data/flowers/`.

[36] P. Welinder, S. Branson, T. Mita, C. Wah, F. Schroff, S. Belongie, and P. Perona, "Caltech-UCSD Birds 200 dataset, California Institute of Technology. CNS-TR-2010-001. 2010.," Available online at: `http://www.vision.caltech.edu/visipedia/CUB-200.html`.

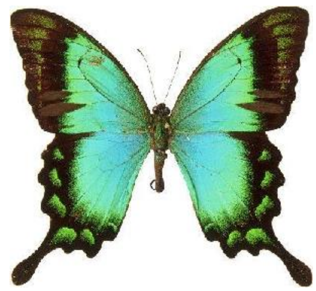

Fig. 2. Example of a query image.

(a)

(b)

Fig. 3. Outcomes of our image-to-image querying system: (a) examples of retrieved images; (b) examples of rejected images.