An Improvement of Dissimilarity-Based Classifications Using SIFT Algorithm*

Evensen E. Masaki and Sang-Woon Kim

Dept. of Computer Science and Engineering,
Myongji University, Yongin, 449-728 South Korea
{evenalexjr, kimsww}@mj.ac.kr

Abstract. In dissimilarity-based classifications (DBC), classifiers are not based on the feature measurements of individual objects, but rather on a suitable dissimilarity measure among the objects. In this paper, we study a new way of measuring the dissimilarity between two object images using a SIFT (Scale Invariant Feature Transformation) algorithm [5], which transforms image data into scale-invariant coordinates relative to local features based on the statistics of gray values in scale-space. With this method, we find an optimal or nearly optimal matching among differing images in scaling and rotation, which leads us to obtain dissimilarity representation after matching them. Our experimental results, obtained with well-known benchmark databases, demonstrate that the proposed mechanism works well and, compared with the previous approaches, achieves further improved results in terms of classification accuracy.

1 Introduction

Dissimilarity-based classifications (DBC) [6] are a way of defining classifiers among classes; the process is not based on the feature measurements of individual objects, but rather on a suitable dissimilarity measure among them. The problem with this strategy is that we need to measure the inter-pattern dissimilarities for all the training samples to ensure there is no zero distance between objects of different classes. Several strategies can be employed to address this problem, which include the regional distance [1], the Euclidean distances of images [9], a dynamic programming technique [4], and classification of regions of interest (ROIs) [8]. In [9], Wang et al. propose a new Euclidean distance, called IMage Euclidean Distance (IMED). Unlike the traditional one, IMED takes into consideration the spatial relationships of pixels. From this point of view, the researchers claim to get a more reliable similarity measure.

In addition, in [4], Kim and Gao propose a way of measuring the dissimilarity distance between two images of an object when the images have different directions and sizes and there is no direct feature correspondence. In this method, a dynamic programming technique, such as dynamic time warping, is used to overcome the limitations of one-to-one mapping. Recently, in [8], Sorensen et al. proposed a method of directly measuring dissimilarities between images by using region of interests (ROIs) of the segments. When computing the dissimilarity between two images, x_1 and x_2 , they consider

* This work was supported by the National Research Foundation of Korea funded by the Korean Government (NRF-2011-0002517).

an associated weight, $\Delta(x_{1i}, x_{2j})$, that expresses the textural dissimilarity between the two corresponding ROIs, x_{1i} and x_{2j} . Here, to avoid a case in which all ROIs in x_1 are matched to all ROIs in x_2 , they employ a way of matching every ROI in one image with the most similar ROI in the other image.

On the other hand, Lowe [5] proposed the scale invariant feature transformation (SIFT) algorithm for extracting distinctive invariant features from images that can be used to perform reliable matching between different views of an object. The features are invariant to image scale and rotation, and are shown to provide robust matching across a substantial range of affine distortion, addition of noise, and change in illumination. Thus, SIFT can be used to detect distinct features in an image and to transform image data into scale-invariant coordinates relative to local features. The algorithm consists of four main steps to generate the set of image features: scale-space peak detection, keypoint localization, orientation assignment, and keypoint descriptor.

The major task of this study is to deal with how dissimilarity measure can be effectively computed. However, when there are many kinds of variations based on such factors as pose (direction), scaling, and illumination [1], [3], it is difficult to improve the performance of DBCs in the dissimilarity space. To overcome this limitation and thereby improve the classification performance of DBCs, in this paper, we study a new way of enriching the representational capability of dissimilarity measures. In particular, this goal can be achieved by using a SIFT algorithm based on the statistics of gray values in scale-space. With regard to measuring the dissimilarity of the object images, we prefer not to directly measure the dissimilarity from the images; rather, we employ a particular way of using SIFT to *adjust* or *scale* the object images. This measure effectively serves as a new “feature” component in the dissimilarity space.

2 DBCs with SIFT Algorithm

Scale Invariant Feature Transformation (SIFT) [5]: SIFT is an algorithm for extracting distinctive invariant features from images that can be used to perform reliable matching between different views of an object. The features are invariant to image scale and rotation, and are shown to provide robust matching across a substantial range of affine distortion, changes in 3D viewpoint, addition of noise, and changes in illumination [5]. Four computational stages of SIFT are scale-space peak detection, keypoint localization, orientation assignment, and keypoint descriptor, respectively. The details of SIFT are omitted here in the interest of brevity, but can be found in the literature [5].

Dissimilarity-Based Classifications (DBC) [6]: A dissimilarity representation of a set of samples, $T = \{x_i\}_{i=1}^n \in \mathbb{R}^d$, is based on pairwise comparisons and is expressed, for example, as an $n \times m$ dissimilarity matrix $D_{T,Y}[i, j]$, where $Y = \{y_j\}_{j=1}^m \in \mathbb{R}^d$, a prototype set, is extracted from T , and the subscripts of D represent the set of elements, on which the dissimilarities are evaluated. Thus, each entry, $D_{T,Y}[i, j]$, corresponds to the dissimilarity between the pairs of objects, x_i and y_j . Consequently, an object, x , is represented as a column vector, $\delta(x)$. In image classification, one of the most intractable problems is the distortion caused by the differences in illumination, scaling, and rotation. To overcome this problem, in this paper, we use the SIFT algorithm. The proposed approach, which is referred to as DBC-with-SIFT, is summarized as follows:

1. Select the whole training set, T , as the representation subset, Y .

2. For all the pairs of objects, x_i and y_j , where $x_i \in T$ and $y_j \in Y$, compute $D_{T,Y}[i, j]$ as follows: (1) After extracting the entire set of *keypoints* from x_i and y_j with SIFT, select the primary *keypoints* among them using heuristic methods¹. (2) Compute $D_{T,Y}[i, j]$ in the following sub-steps: The first sub-step is to generate two queues, Q_x and Q_y , by inserting the selected primary keypoints and $d = 0$. The second sub-step is to compute the Euclidean distance, $d = d + \|x(i, j) - y(k, l)\|$, between two pixels (keypoints), $(i, j) \in Q_x$ and $(k, l) \in Q_y$, which have been deleted from the two queues, respectively. Then, the 8-nearest neighbors around the pixels are inserted into the queues. If Q_x and Q_y are empty, then exit; otherwise, go to the second sub-step.

3. For a testing sample, z , compute $\delta(z)$ by using the same measure used in Step 2.

4. Achieve the classification by invoking a classifier built in the dissimilarity space defined with $D_{T,Y}[\cdot, \cdot]$ and by operating the classifier on the dissimilarity vector, $\delta(z)$.

The time complexities (i.e., the increase in computational complexity) of the above algorithm are omitted here, but will be presented in the journal version.

Primary Keypoint Selections: In DBC-with-SIFT, to select the primary keypoints to compensate for the differences in direction or illumination, we consider two heuristic rules: *slope-based (SB)* rule and *number-based (NB)* rule. For each pair of keypoints, slope values, $\{\phi_i\}_{i=1}^M$, where ϕ_i is the angle between the i^{th} slope and a base line and M the number of the keypoints, can be calculated by using their frames². In the SB rule, after computing all the slope values of the keypoints, we select *primary* keypoints from the keypoints, which are within the range of $-\alpha \leq \phi_i \leq \alpha$, where $\{\alpha | \alpha = 0.2, 0.3, \dots, 1.0\}$. In the NB rule, on the other hand, after sorting all the slope values, we select the β number of keypoints in an ascending order from the smallest values of $\{\phi_i\}_{i=1}^M$ as the primary keypoints, where $\{\beta | \beta = 1, 3, \dots, 17\}$.

3 Experimental Results

Experimental Data: The proposed method has been tested and compared with conventional methods. This was done by performing experiments on three benchmark databases, namely, AT&T³, CMU (CMU-Expression)⁴, and Leaves⁵. Here, the numbers of samples, dimensions, and classes of these databases are, respectively, (400, 10304, 40), (200, 16384, 10), and (45, 530432, 3). Also, Leaves was employed as a benchmark database for object classification, while AT&T and CMU were utilized as examples of the non-background and background face recognition, respectively.

Experimental Method: First, data sets are split into training sets and test sets in the ratio of 75 : 25. The training and testing procedures are repeated 30 times and the results obtained are averaged. Next, in the conventional DBCs, which is referred to as

¹ This method will be presented in a later section.

² The SB rule is based on assumption that face images are well posed in vertical shift and size. If such positioning is not perfect, some special detectors of human face could be considered.

³ <http://www.cl.cam.ac.uk/Research/DTG/attarchive/facedatabase.html>

⁴ http://www.ri.cmu.edu/research_project_detail.html?project_id=418&menu_id=261

⁵ <http://www.vision.caltech.edu/html-files/archive.html>

Fig. 1. Plots of the pairwise keypoints obtained with SIFT and the primary keypoints selected with the NB rule for two objects of AT&T: (a) left and (b) right; (a) is for the entire set of keypoints. (b) is for the primary keypoints selected from (a).

DBC_s-without-SIFT (shortly SIFT-without), we computed the dissimilarity between two object images using their gray values. In DBC_s-with-SIFT (shortly SIFT-with), however, after extracting the primary keypoints from the two objects using SIFT and NB (or SB) rule, we computed the dissimilarity. For example, Fig. 1 shows plots of the pairwise keypoints and the selected primary keypoints for two objects of AT&T. Finally, to evaluate DBC_s built in the SIFT-without and SIFT-with methods, different classifiers, such as k -nearest neighbor classifiers (referred to *nnc*) and support vector machines (named *svm*), are employed and implemented with PRTTools [2].

Experimental Results: Fig. 2 shows a comparison of the estimated error rates of *two* classifiers designed with the SIFT-without (*nnc1* and *svm1* - dashed lines of $\circ$ and $\triangleleft$ markers, respectively) and SIFT-with (*nnc2* and *svm2* - solid lines of the same markers) methods for the databases.

First, the estimated error rates for AT&T are very small but for the rest of the databases (i.e., CMU and Leaves) are higher. This might be because when capturing photos, the former method is non-background. Next, from the pictures shown in Fig. 2, it should be observed that, in general, the dashed lines are higher than the solid lines. That is, for the three databases, the estimated error rates of *nnc1* and *svm1* are higher than those of *nnc2* and *svm2*, respectively. From this observation, the following conclusion is derived: measuring dissimilarities between images with SIFT-with method leads to lower error rates than measuring the distance with SIFT-without method.

The other observations obtained from the figures are the following:

- For AT&T (SB), the *svm*-lines of $\triangleleft$ markers show that there is more improvement in classification error rate when using the SB rule than when using the NB rule. This indicates that we have achieved the goal that we desired to solve using the proposed methods (refer to Figs. 2(a) and (b) on top).

- For AT&T and Leaves, it should be observed that *svm* is the best classifier for the SB and NB rules in terms of classification accuracies. However, for CMU database, *svm* is the best classifier for SB rule, while for the case of NB rule, the best working classifier varies: For $1 \leq \beta \leq 5$ and $14 \leq \beta \leq 17$, *svm* works well. For $6 \leq \beta \leq 13$, however, *nnc* acts well (refer to Figs. 2(c) and (d) in middle).

- For the Leaves database, SB rule gives almost the same estimated error rates (flat), but for the case of NB rule, there is an increase and decrease of estimated error rates, which makes it difficult to find the optimal or nearly optimal slope values and the number of keypoints (refer to Figs. 2(e) and (f) at bottom).

Fig. 2. A comparison of the estimated error rates of *two* classifiers designed with the SIFT-without (*nnc1* and *svm1* - dashed lines of $\circ$ and $\triangleleft$ markers) and SIFT-with methods (*nnc2* and *svm2* - solid lines of the same markers) for the three databases: (a) top left, (b) top right, (c) middle left, (d) middle right, (e) bottom left, and (f) bottom right; (a) and (b) are for the two rules (i.e., SB and NB rules) for AT&T; (c) and (d) are for the two methods of CMU; (e) and (f) are for Leaves

4 Conclusions

In order to improve the classification performance of DBCs, in this paper we used the SIFT algorithm based on the local statistics of gray values in scale-space. With SIFT, we first obtained the entire set of keypoints from paired images and, in turn, selected a few primary keypoints using either the slope-based or the number-based rule, eliminating inappropriate keypoints. Using the selected primary keypoints, we generated a dissimilarity representation, in which DBCs have been performed. Our experimental results, obtained with three benchmark databases, demonstrate that the classification accuracy of DBCs was improved when the number of primary matched keypoints was appropriately chosen. Although we have shown that DBCs can be improved by employing SIFT, many tasks remain. One of them is to improve the classification efficiency by selecting an optimal or nearly optimal number of the primary keypoints from the entire set of keypoints obtained with SIFT for the paired object images. Also, it is not yet clear which kinds of significant data sets are more suitable for the scheme.

References

1. Adini, Y., Moses, Y., Ullman, S.: Face recognition: the problem of compensating for changes in illumination direction. *IEEE Trans. Pattern Anal. and Machine Intell.* 19(7), 721–732 (1997)
2. Duin, R.P.W., Juszczak, P., de Ridder, D., Paclik, P., Pekalska, E., Tax, D.M.J.: *PRTTools 4: a Matlab Toolbox for Pattern Recognition*. Technical Report, Delft University of Technology, The Netherlands (2004), <http://prtools.org/>
3. Georgiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. and Machine Intell.* 23(6), 643–660 (2001)
4. Kim, S.-W., Gao, J.: A dynamic programming technique for optimizing dissimilarity-based classifiers. In: da Vitoria Lobo, N., Kasparis, T., Roli, F., Kwok, J.T., Georgiopoulos, M., Anagnostopoulos, G.C., Loog, M. (eds.) *S+SSPR 2008*. LNCS, vol. 5342, pp. 654–663. Springer, Heidelberg (2008)
5. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
6. Pekalska, E., Duin, R.P.W.: *The Dissimilarity Representation for Pattern Recognition: Foundations and Applications*. World Scientific Publishing, Singapore (2005)
7. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression (PIE) database of human faces. Technical report CMU-RI-TR-01-02, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA (2001)
8. Sørensen, L., Loog, M., Lo, P., Ashraf, H., Dirksen, A., Duin, R.P.W., de Bruijne, M.: Image dissimilarity-based quantification of lung disease from CT. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6361, pp. 37–44. Springer, Heidelberg (2010)
9. Wang, L., Zhang, Y., Feng, J.: On the Euclidean distance of images. *IEEE Trans. Pattern Anal. and Machine Intell.* 27(8), 1334–1339 (2005)