Integration of Local and Global Shape Analysis for Logo Classification

Jan Neumann[a], Hanan Samet[a*], Aya Soffer[a†]

[a]Computer Science Department
Center for Automation Research
Institute for Advanced Computer Studies
University of Maryland
College Park, Maryland 20742
jneumann@cfar.umd.edu, hjs@cs.umd.edu, ayas@il.ibm.com

A comparison is made of global and local methods for the shape analysis of logos in an image database. The qualities of the methods are judged by using the shape signatures to define a similarity metric on the logos. As representatives for the two classes of methods, we use the negative shape method which is based on local shape information and a wavelet-based method which makes use of global information. We apply both methods to images with different kinds of degradations and examine how a particular degradation highlights the strengths and shortcomings of each method. Finally, we use these results to develop a new adaptive weighting scheme which is based on the relative performances of the two methods. This scheme gives rise to a new method that is much more robust with respect to all degradations examined and works by automatically predicting if the negative shape or wavelet method is performing better.
Keywords: shape representation, shape recognition, image databases, symbol recognition, logos

1. Introduction

We examine three different approaches for classifying images with several components in an image database. One approach uses a local method to represent the image, the second uses a global method, while the third combines both using an adaptive weighting scheme based on relative performance. The local method uses so-called negative symbols, as described in [7], to compute a number of statistical and morphological shape features for each connected component of an image foreground and background. The global method uses a wavelet decomposition of the horizontal and vertical projections of the global image as described in [3]. As a sample application of well-defined multi-component images, we use logos. Several studies have reported results on some form of logo recognition. Each study used either global or local methods. These include local invariants [2,6], wavelet features [3], neural networks [1], and graphical distribution features [5]. The performance

* The support of the National Science Foundation under Grants CDA-95-03994, IRI-97-12715, EIA-99-00268, and IIS-00-86162 is gratefully acknowledged.
† Currently at IBM Research Lab, Haifa 31905, Israel.

in case of certain degradations was examined. In this paper we compare the local and global methods under the influence of several image degradations. The performance measure is the ranking of the original logo after inputing a degraded version of it into the classifier. The results exhibit the advantages and disadvantages of local methods, based on shape features, in contrast to global methods, rooted in signal processing. Finally, we present an algorithm that combines both methods into a single, robust framework by adaptively weighting the contributions of each method according to an estimate of their relative performance.

2. Wavelet Method

We studied logos in the UMD-Logo-Database which are gray-scale images that are scanned versions of black and white logos (Figure 1a). We assume that the logos have already been segmented and binarized by a preprocessing step. The problems of segmentation and threshold selection are beyond the scope of this paper. The classification methods should be scale, translation, and rotation invariant. To achieve this, we apply some normalizing steps to the input images before we start the computation of any features (Figure 1b). After shifting the image so that the centroid of the white pixel locations is located at the image center, which gives us translational invariance, we rotate the image around the centroid so that its major principal axis is aligned with the horizontal. This gives us rotational invariance. Finally, we resize the image so that the bounding box of the logo symbol is a given percentage of the image size. This accounts for changes in scale of the input logos. These transformations make it possible to perform the following computations without reference to orientation, position, and scale.

Given a normalized image we compute the horizontal and vertical projections (Figures 1c–1d) of this binary image which are defined as $P(y) = \sum_{x=1}^{n} I(x, y)$ and $P(x) = \sum_{y=1}^{m} I(x, y)$. This means that we are counting the number of white pixels for each row and column. Next, we use a wavelet transform to low-pass filter the projections (Figures 1e–1f). In our experiments, we used the Haar wavelet and the Daubechies wavelet. This process is illustrated in Figure 1 for the Haar wavelet. These coefficient vectors, called *signatures*, are used to compare different logos among each other. We use the L_1-Norm to compute the difference between their signatures because the L_1-Norm is known to be robust against outliers [8].

3. The Negative Shape Method

The novel idea of the negative shape method as defined in [7] for the representation of symbol-like data such as found in logos is that we compute the shape features not just for the components of the foreground that constitute the symbol itself, but also for the components that make up the background of the image containing the symbol.

We start by labeling the connected components of the image and its background. To each component of the labeled image, we first apply the preprocessing steps described in Section 2 and then compute the following shape features:

1. **F1: Invariant moment:** Normalized distance of pixel positions to the centroid.
2. **F2: Eccentricity:** The ratio between the lengths of the major and minor axes.

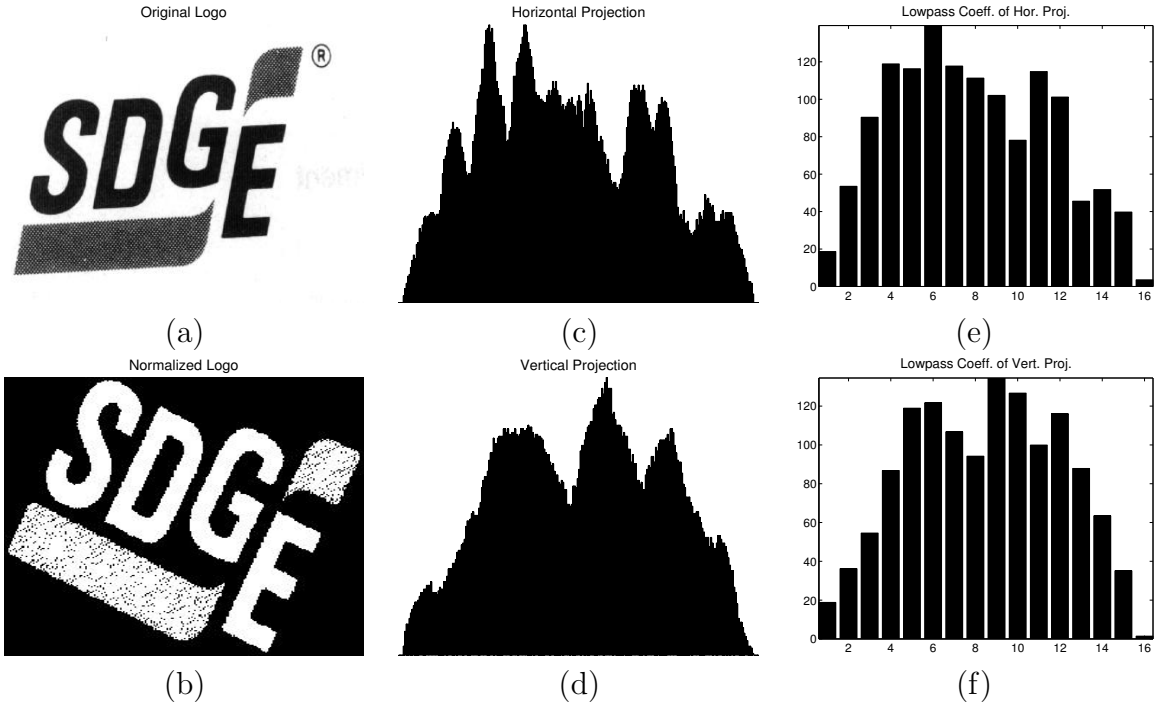


Figure 1. The steps of the wavelet method: (a) original image, (b) normalized image, (c) horizontal projection, (d) vertical projection, (e) low-pass wavelet coefficients of the horizontal projection, (f) low-pass wavelet coefficients of the vertical projection (x-axis: index of coefficient, y-axis: coefficient magnitude).

3. **F3: Circularity:** The ratio between the perimeter of the component and the perimeter of a circle of equivalent area.

4. **F4: Rectangularity:** The ratio between the area of the component and the area of its bounding box.

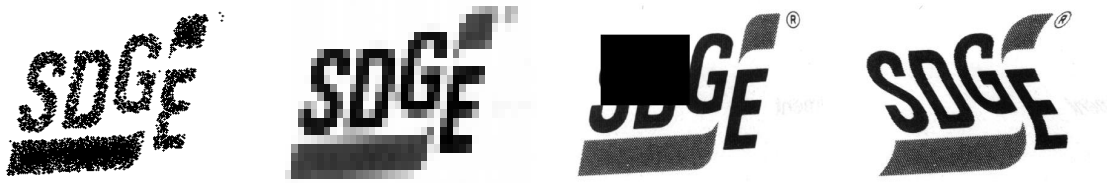
5. **F5: Hole Area Ratio:** The ratio between the area of the holes inside the component and the area of the solid part of the component.

6. **F6,F7: Horizontal (Vertical) Gap Ratio:** The square of the number of pixels inside the component that have a right (bottom) neighbor that does not belong to the component divided by the area of the component.

For the negative shape method we define the distance measure between two logos L_1 and L_2 as follows. First, we normalize the value range for each element of the feature vector over all the logos in the dataset. Then, for each component of L_1 we find the component of L_2 that has the smallest Euclidean distance in feature space to it. Finally, the average of these minimal distances over all the components of L_1 yields a measure for the distance between the two logos.

4. Comparison Between the Methods

All methods were implemented in Matlab^{TM} and were applied to the logos contained in the UMD-Logo-Database (123 logos, can be found at [9]). The system was tested by providing it with an input logo and ranking the logos in the database based on their similarity to this logo. Below, we investigate the robustness of the methods when the logos are corrupted using four different image degradation methods. We choose two global degradations as demonstrated in Figures 2a and 2b, and two local degradations as shown in Figures 2c and 2d which we will describe in more detail later. Global degradation means that the degradation influences all regions of the image in the same way while leaving the large-scale image structure intact. Local degradations either influence only a localized image region (Figure 2c) or leave the small-scale image structure invariant (Figure 2d).



(a) Sensor Noise (b) Subsampling (c) Occlusion (d) Swirl

Figure 2. The two global (a–b) and two local (c–d) types of degradations.

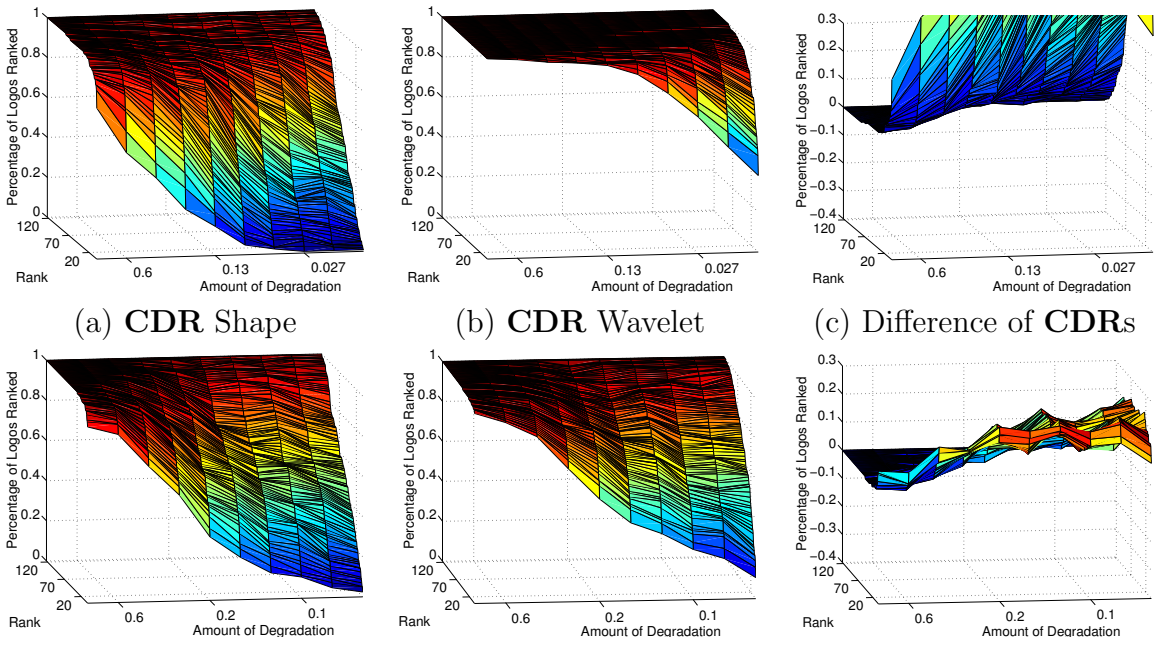
For each method, we degrade the images in the database, varying the amount of degradation in ten equally spaced steps. We then classify the logos in the database based on the computed distance in the feature space to the degraded input, before we examine the rank–in terms of feature space distance–of the corresponding uncompromised logo. To analyze the influence of a degradation on the performance of a method, we plot the cumulative distribution of the ranks of the original logos which is defined as follows. Let i be the index of a logo and n the number of logos in the database. If we degrade logo i and input it into the classifier, the corresponding original, uncompromised logo will have rank $rk(i)$. The cumulative distribution of the ranks (**CDR**) is now defined as

$$\mathbf{CDR}(R) = \frac{1}{n}\sum_{i=1}^{n}\phi(rk(i),R) \text{ where } \phi(rk(i),R) = \begin{cases} 1 & : \quad rk(i) \leq R \\ 0 & : \quad rk(i) > R \end{cases}. \tag{1}$$

This monotone function indicates the percentage of the logos for which the classifier has ranked their undegraded match in the database among the first R logos thereby giving us a means to measure the performance of each classifier. In order to show the difference in performance between the methods that we used, we also included graphs displaying the difference of the respective **CDR**s. Note that the closer the **CDR** is to 1, the better the classifier performs because the correct matches have been given high ranks.

4.1. Global Degradations: Optical Sensor Noise and Reduced Resolution

To simulate the image degradation that is caused by processes such as fax transmissions or photo copying, we use the well-known document degradation model by Kanungo, Haralick, and Baird [4]. It models the inversion of individual pixel (from foreground to background and vice versa) due to light intensity fluctuations, the sensitivity of the sensor and the thresholding level used in the binarization, as well as the blurring caused by the point-spread function of the scanner optical system. We first computed the distance d of each pixel from the logo boundary, then we flipped each pixel value with the probability $p(0|1, d, \alpha) = p(1|0, d, \alpha, \alpha_0) = \alpha_0 e^{-\alpha d^2}$. Finally, we correlated the value of neighbouring pixels by performing a morphological closing operation with a disk structuring element of diameter k. For more details, see [4]. In our experiments, we set $\alpha_0 = 0.5$ and $k = 2$, and α was varied logarithmic steps from 0.05 to 1 as indicated on the x-axis in Figures 3a–3c. An example degraded image is given in Figure 2a.



(a) **CDR** Shape (b) **CDR** Wavelet (c) Difference of **CDR**s

(d) **CDR** Shape (e) **CDR** Wavelet (f) Difference of **CDR**s

Figure 3. Performance under global degradations: (a–c) show the results when we added optical sensor noise and (d–f) when we reduced the resolution. We see in (a) and (d) that the shape method is very sensitive to global degradations, and is noticeably outperformed by the wavelet method as seen in the difference plots in (c) and (f), where we subtract the **CDR** of the negative shape method from the **CDR** of the wavelet method.

The wavelet method outperforms the negative shape method noticeably as can be seen in Figure 3c. Even for large amounts of noise, the wavelet method classifies accurately as indicated by the flat **CDR** in Figure 3b, while the performance of the negative shape method worsens rapidly (Figure 3a). It is to be expected that the wavelet method outper-

forms the negative shape method when we simulate optical sensor noise because flipping foreground and background pixels with equal probability should have only a small effect on the distribution of white and black pixels in a row or column, and thus the projection should not change much. On the other hand, in the negative shape method, we compute the feature vectors only on a small subset of the pixels of each component. In this case, the noise will change the spatial distribution of the pixels more drastically because of the smaller number of pixels involved. Thus the negative shape method is less robust towards optical sensor noise than the wavelet method.

To see how the methods handle differences in image resolution, which is obviously not offset by the scaling invariance since we work on digitized images, we reduce the size of the input images through sub-sampling using bilinear interpolation (e.g., Figure 2b). The parameter value is the size ratio between the original and the sub-sampled image as indicated on the x-axis in Figures 3d–3f).

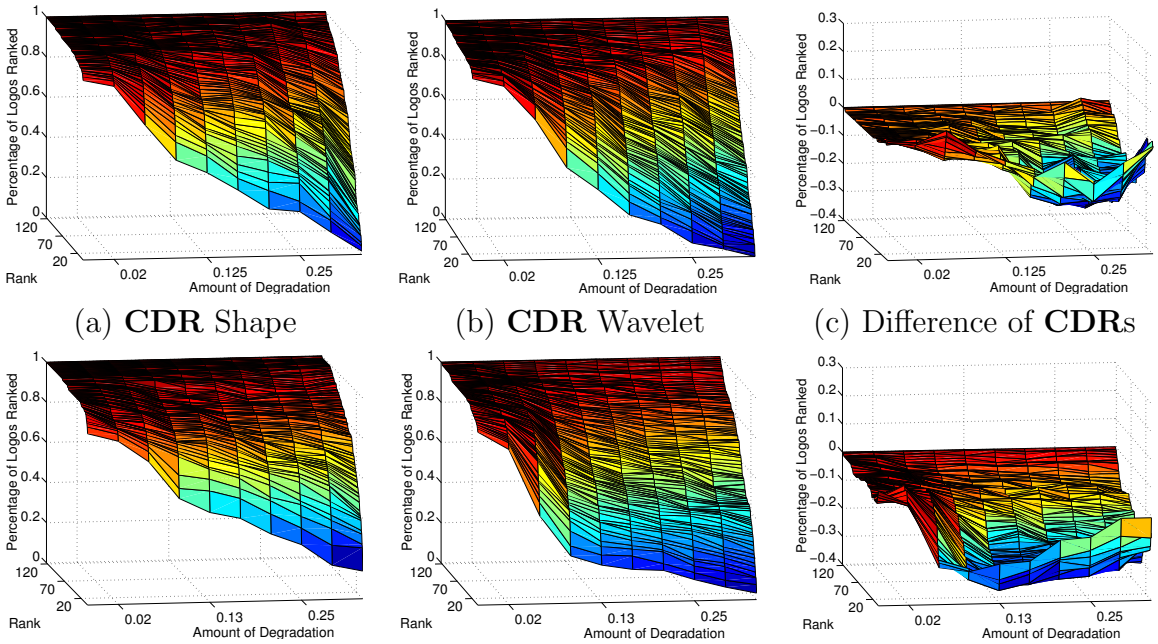
As before, the wavelet method outperforms the negative shape method (Figure 3f), although the negative shape method does not exhibit the same breakdown in performance as in the case of random noise (Figure 3d). Since we use the low-pass wavelet coefficients for the classifier, the reduced resolution does not influence the performance of the wavelet method drastically. This is because subsampling an image by bilinear interpolation has a similar effect as low-pass filtering the image. The low-pass wavelet coefficients of a low-pass filtered image are in general very similar to the low-pass coefficients computed on the original image. As before, the negative shape method is affected by this degradation because even when large scale changes are hardly visible, local shape features such as circularity, rectangularity and gap ratios are more susceptible to local changes due to a loss of detail.

4.2. Local Degradations: Occlusion and Swirl

To model the occlusion of parts of a logo, we add a component to the logo image which in this case is a black rectangle of varying size (see Figure 2c). The parameter here is the percentage of the image that is occluded by the rectangle.

The performance graphs show that occlusion has a greater effect on the wavelet method than the negative shape method (Figures 4a–4c) although both methods are able to handle small occlusions well. Since the addition of an extra object or the omission of parts of the image causes global changes to the distribution of pixels in each row or column, the projections are strongly affected and thus so are the wavelet coefficients. Because of the local structure of the shape features, the components that are not occluded are not degraded at all and their feature values are unchanged. In the classifier we average the best feature vector matches for all the components in the input image. Since an occlusion is more likely to combine components into larger aggregates than to break them into many new ones, these few new components which do not have a corresponding component in the original image, are influencing the feature distance only to a small extent. Except for very degenerate configurations, the influence of the new components is averaged out by the continuing good matches of the feature vectors of the remaining uninfluenced components.

Swirling is a smooth deformation of an image which can be used to model a non-isotropic stretching of a logo. The relative position of each row is shifted to the left or right by an offset given by a smooth function, where the offset is limited to a certain percentage



(a) **CDR** Shape (b) **CDR** Wavelet (c) Difference of **CDR**s

(d) **CDR** Shape (e) **CDR** Wavelet (f) Difference of **CDR**s

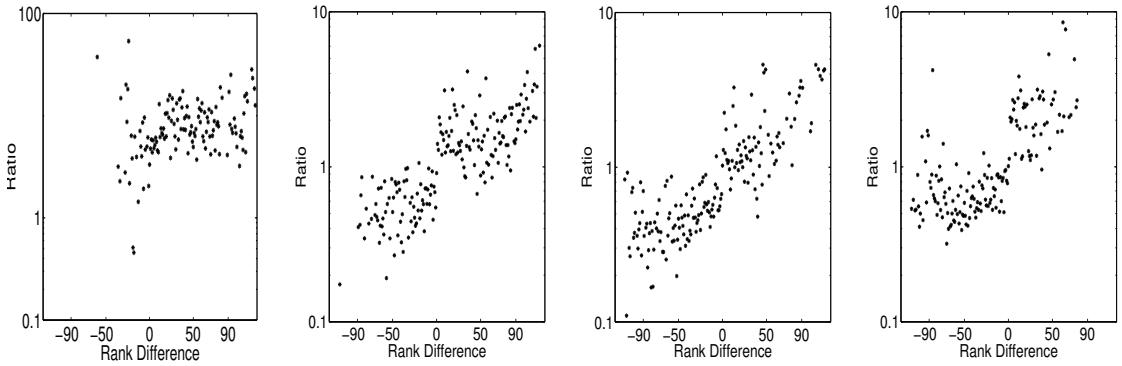
Figure 4. Local degradations: (a–c) shows the results for the occlusion and (d–f) the results for the swirl. Notice how the local degradations affect the performance of the wavelet method much more than the negative shape method as indicated in the difference plots (c) and (f), where we subtract the **CDR** of the negative shape method from the **CDR** of the wavelet method.

of the image width which is given as a parameter. This deforms the logo as if we would stretch a rubber sheet in different directions (see Figure 2d).

This degradation has very different effects on the two methods. The performance of the wavelet method worsens rapidly with increasing swirl until we basically get a random ranking, as indicated by the linear increase of the CDR in Figure 4e. In contrast, the negative shape method classifies much better as seen in the difference of the **CDR**s in Figure 4f because it is possible to approximate this deformation by localized similarity transforms since the deformation is smooth. The local features used by the negative shape method are invariant to similarity transforms due to the component normalization. Therefore, it is much less affected by this degradation than the wavelet method. Recall that the wavelet method is only globally invariant to similarity transforms due to the global preprocessing, but not locally.

5. The Best of both Worlds: A Combination of the two Methods

In Section 4 we saw that the wavelet and the negative shape methods perform very differently if the input logo is corrupted by either local or global degradations. Thus, if we were able to detect which method is performing better, then we could devise a switching mechanism that chooses the better performing method automatically.



(a) Sensor Noise (b) Subsampling (c) Occlusion (d) Swirl

Figure 5. Relationship between relative feature distance and classification performance. For each type of degradation, we plot the difference D in the classification of the matching logo (for all logos i, we subtract the $rk(i)$ as determined by the wavelet method from $rk(i)$ as determined by the shape method), and compare it to $R_{rankdiff}(D)$, the geometric mean of the ratios as defined in Equation 2.

To be able to compare both methods, we first normalize the featured distances by computing for both methods the average of the pairwise feature distances between all uncompromised logos in the database. We denote these normalization factors by N_{shape} and N_{wave}. For each type of degradation, we vary the amount of degradation in ten equally-spaced steps and compute the following measure over all logos in the database and for all amounts of degradations:

$$R_{rankdiff}(D) := \left(\prod_{i \in \mathcal{L}} \frac{F_{shape}(i)}{N_{shape}} \frac{N_{wave}}{F_{wave}(i)} \right)^{\frac{1}{|\mathcal{L}|}}, \tag{2}$$
$$\text{where } \mathcal{L} := \{i | rk_{shape}(i) - rk_{wave}(i) = D\}.$$

$F_{shape}(i)$ is the feature distance between a degraded version of logo i and the undegraded version of i as computed by the negative shape method and $F_{wave}(i)$ the feature distance as computed by the wavelet method, while $rk_{shape}(i)$ and $rk_{wave}(i)$ are the corresponding ranks of the matching undegraded logo (see Section 4). This measure correlates each difference in ranking with the geometric mean of the corresponding ratios between the feature distances computed by the two methods. Therefore, we relate the performance of the classifier with the relative feature space distances. Note that we use the geometric mean instead of the arithmetic mean as $R_{rankdiff}$ involves a product of ratios rather than a sum of values.

Figure 5 shows that whenever one method performs better (positive rank difference if the wavelet method is performing better, negative if the negative shape method performs better), this method in general also computes a lower normalized feature distance for the original logo (the ratio is larger than 1 if the wavelet method computes the lower

normalized feature distance, the ratio is smaller than 1 if the negative shape method computes the lower normalized feature distance). Understanding this relationship between the change in the ratio of the normalized feature distances and the relative performance of the two methods when applied to degraded images enabled us to adaptively weight the respective contributions of the two methods when combining them into a single distance measure. Therefore, in order to take advantage of the respective strengths of both methods we devised the following performance-dependent weighting scheme.

Let the vector of feature distances of the degraded logo i to all undegraded logos in the database be $D_{shape}(i)$ for the negative shape method and be $D_{wave}(i)$ for the wavelet method. Moreover, denote their means by $A_{shape(i)}$ and by $A_{wave}(i)$, respectively. Now, if the ratio

$$R_{an}(i) = \frac{A_{shape}(i)}{N_{shape}} \frac{N_{wave}}{A_{wave}(i)} \tag{3}$$

indicates that one of the two methods is performing worse than expected, then we decrease its weight in the final classification and increase the weight of the other method. The combined feature distance vector $D_{comb}(i)$ for the degraded input logo i is then a weighted sum of $D_{wave}(i)$ and $D_{shape}(i)$:

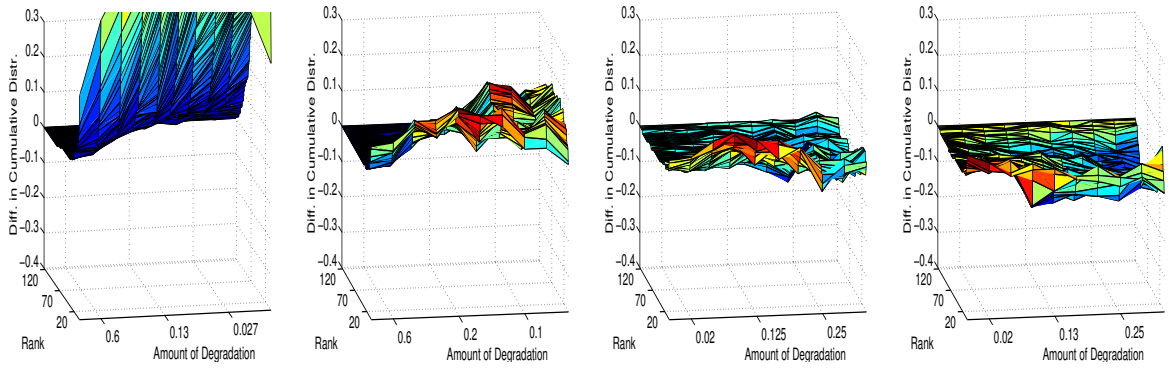
$$D_{comb}(i) = R_{an}(i) \frac{D_{wave}(i)}{N_{wave}} + R_{an}(i)^{-1} \frac{D_{shape}(i)}{N_{shape}} \tag{4}$$

If the ratio R_{an} decreases (increases), this indicates that the degradation of the input logo causes the wavelet method to compute feature distances larger (smaller) than the negative shape method. Consequently, the weight of D_{wave} with respect to D_{shape} in Equation 4 will be reduced (increased), because we have less (more) confidence in the wavelet method's ability to classify the input logo correctly than in the negative shape method's ability to do so.

This adaptive weighting scheme increases the robustness of the classification dramatically. Comparing the performance of the combined method on the images degraded as described in Section 4, we see that our weighting scheme is able to detect the change in relative performance and adjust the weights to mimic the classification of the better performing method. Almost always the combined method equals the performance of the better method (Figures 6c,d,e,f) and surpasses by far the worse method (Figures 6a,b,g,h) in terms of the **CDR**. This shows that our combined scheme is effective in capturing global as well as local shape information and is thus able to deal well with all the image degradations of the kind that we described.

6. Summary and Future Work

Both the wavelet as well as the negative shape method are well-suited for certain kinds of image degradations but are very sensitive to others. This discrepancy in performance can be explained by the difference between local shape feature-based methods and global



(a) Noise (b) Subsampling (c) Occlusion (d) Swirl

(e) Noise (f) Subsampling (g) Occlusion (h) Swirl

Figure 6. Performance of the combined method: in (a)–(d) we subtract the **CDR** of the negative shape method from the **CDR** of the combined method, and in (e)–(h) we subtract the **CDR** of the wavelet method from the **CDR** of the combined method.

filter-based methods. On the one hand, we have the wavelet method that operates on the global image and computes features that are relatively invariant to degradations that are isotropic. On the other hand, we have the negative shape method which operates on local image regions. Thus its features are relatively invariant to changes that leave the image at other locations mostly intact such as occlusions or preserve the local image structure such as the swirl deformation. We take advantage of the fact that both basic methods perform very differently on images that exhibit degradations of either local or global nature by devising a performance-dependent weighting scheme that combines the results of both methods. Our combined algorithm shows a noticeable improvement in the robustness of the classification by combining the strengths and avoiding the weaknesses of the respective methods. This weighting scheme performs the better the more different the performances of the underlying methods are because this makes it easier to detect if one method is performing poorly with respect to the other method. Therefore, the wavelet and the negative shape methods are very well-suited to be combined by a performance-dependent weighting scheme. Of course, the overall performance of our adaptive scheme is limited by the best performance of either method, thus the application of this weighting scheme to more advanced logo recognition methods is planned for future work. In addition, we will

try to improve the synergy between the two methods by using local image information to estimate how much an image region is degraded and then use this locality information to adaptively weight the feature vectors on the component level.

REFERENCES

1. F. Cesarini, E. Fracesconi, M. Gori, S. Marinai, J. Q. Sheng, and G. Soda. A neural-based architecture for spot-noisy logo recognition. In *Proceedings of the Fourth International Conference on Document Analysis and Recognition*, pages 175–179, Ulm, Germany, August 1997.
2. D. Doermann, E. Rivlin, and I. Weiss. Applying algebraic and differential invariants for logo recognition. *Machine Vision and Applications*, 9(2):73–86, 1996.
3. M. Y. Jaisimha. Wavelet features for similarity based retrieval of logo images. In *Proceedings of the SPIE, Document Recognition III*, volume 2660, pages 89–100, San Jose, CA, January 1996.
4. T. Kanungo, R. M. Haralick, H. S. Baird, W. Stuezle, and D. Madigan. Statistical, nonparametric methodology for document degradation model validation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1209–1223, November 2000.
5. T. Kato. Database architecture for content-based image retrieval. In *Proceedings of the SPIE, Image Storage and Retrieval Systems*, volume 1662, pages 112–123, San Jose, CA, February 1992.
6. M. Kliot and E. Rivlin. Shape retrieval in pictorial databases via geometric features. Technical Report CIS9701, Technion - IIT, Computer Science Department, Haifa, Israel, 1997.
7. A. Soffer and H. Samet. Using negative shape features for logo similarity matching. In *Proceedings of the 14th International Conference on Pattern Recognition*, pages 571–573, Brisbane, Australia, August 1998.
8. G. Strang and T. Nguyen. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA, 1996.
9. http://documents.cfar.umd.edu/resources/database/umdlogo.html.