Performance Evaluation of the Nearest Feature Line Method in Image Classification and Retrieval

Stan Z. Li*

Microsoft Research China

Sigma Center, Beijing 100080, China

Kap Luk Chan Changliang Wang

School of Electrical and Electronic Engineering

Nanyang Technological University, Singapore 639798

Corresponding author: Stan Z. Li, szli@microsoft.com, http://research.microsoft.com/~szli/

Abstract

A new method, the *nearest feature line (NFL)* method, is used in image classification and retrieval, and its performance is evaluated and compared with other methods by extensive experiments. The NFL method is demonstrated to make efficient use of knowledge about multiple prototypes of a class to represent that class.

Keywords

Image classification, image retrieval, nearest feature line (NFL), nearest neighbor (NN) search, similarity metrics.

I. Introduction

Image retrieval finds similar images in the ascending order of similarity or distance, while image classification classifies a query image into the pre-defined classes associated with the top matched image. Both requires a definition of metric to measure the similarity in terms of some distance between images, where a distance is defined based on some images features. Two issues are central to this: (i) what features to use to represent an image, and (ii) how to measure the distance between the images given the chosen representation. This work focuses on the second issue.

Various distance metrics have been used for pattern analysis: Euclidean distance, Cosine distance, Histogram intersection, Hamming distance, Quadratic distance and Mahalanobis distance. A notable commonality among these is that they are defined between the query and an individual prototype. For this reason, the search sees a class as consisting of isolated points in the feature space. There is no class membership concept for the prototypes. The result is that images are sorted in the ascending order of the distances from the query. We refer to this type of search collectively as the nearest neighbor (NN) search, though the NN is originally meant to be a rule for classification [1]. However, in many cases, multiple prototypes are available within an image class. Such a characteristics can be used to improve the classification and retrieval performance but has been ignored by the NN type of methods.

In this paper, we present the use of a new method, called the nearest feature line (NFL) method [2], for image classification and retrieval (Section II), aiming to circumvent the above mentioned limitations of the NN. The assumptions made in the NFL are (i) that the prototypes have been classified into classes *a priori* through some viable means, and (ii) that multiple prototypes are available per class. A novel distance metric is defined to take advantage of these assumptions to improve the performance, with an image representation based on Gabor and wavelet features [3], [4], [5].

In contrast to the NN type metrics, the NFL metric makes use of the available information about classes contained in the multiple prototypes of each class. A subspace is constructed out from the whole feature space for each image class, based on the prior knowledge of multiple prototypes to represent the class. The NFL metric is defined as the Euclidean distance between the query and its projection in the subspace representing the class. The within-class prototypes are generalized to represent variants of that class and thus the generalization ability of the classifier is increased.

An extensive experimental evaluation is carried out to substantiate the strength of the NFL by comparing with other conventional methods, using the Brodatz texture database [6] and a general color image database. Experimental evaluation (Section III) shows that the NFL produces consistently superior results over the NN-type search. For example, the NFL achieves 90.00% retrieval efficiency as opposed to 74.37% of the NN search as reported in Manjunath and Ma [4] for the Brodatz database (when top 15 matches are considered). For the color image database, the NFL has a 75% of increase in retrieval efficiency over the NN method (when top 20 matches are considered). Also, the NFL outperforms k-NN and nearest center (NC) methods, the latter two also using constraints from multiple prototypes per class. This shows that the NFL presents an effective means of using such constraints. The demonstrations can be accessed online at `http://research.microsoft.com/~szli/demos.html`.

II. The Nearest Feature Line (NFL) Method

The rationale of the NFL is based on the following considerations. An image corresponds to a point (vector) in the feature space. When one prototype image changes continuously to another prototype image in some way, it draws a trajectory linking their feature points in the features space. The set of all such trajectories constitute a subspace in the features space representing the class. A similar image should be close to the subspace though may not be so to the original prototypes. The NN search tends to ignore such information.

A. Related Work

The NFL has a close relationship with the linear combination approach [7], the latter being a shape-based approach for recognizing 3D objects from 2D images. It makes use of a linear combination of two prototypes in a feature space, whereas in [7], a 3D object is represented by a linear combination of 2D boundary maps of the object. An object in the image is classified as belonging to a prototypical model object if it can be expressed as a linear combination of the views of the object for some set of coefficients.

A theory of view-based object recognition is presented in [8]. It is based on the observation that the views of a shape-based 3D rigid object undergoing transformation such as rotation reside in a smooth low-dimensional manifold embedded in the space of coordinates of points attached to the object; and for the object, there exists a smooth transformation function which can map any perspective view into another view of the object. Further, it is also demonstrated that this transformation function can be approximated from a small number of views of the object. The theory is further demonstrated in [9] on a variety of objects, and its application is extended from recognition to categorization. However, object recognition in those studies is based on shape information alone; variations in illumination and texture of objects and non-rigid shape changes are not dealt with.

In [10], a technique is presented to synthesize a new image of an object from a single 2D view of the object using a linear combination of images of prototype objects of the same class, with the assumption that the object belongs to *linear object classes*. This approach avoids the use of 3D models for the view synthesis and is capable of generating a new view of a 3D object from a single 2D view of the object, using both shape and texture information. The technique requires correspondence between all feature points of prototype images and between the new image and one of the prototypes.

The feature line on which the NFL is based can be considered as a simpler version of the spline type manifold of the parametric appearance representation [11]. In [11], the appearance manifold of an object is constructed from images of the object taken on a turnable (parameterized by a single parameter) under carefully controlled lighting (parameterized by another single parameter). However, such strictly controlled imaging conditions are difficult to enforce in acquiring images from diverse domains. The NFL does not need such conditions and provides a simple yet useful solution for modeling classes.

The idea of using multiple prototypes has been incorporated in relevance feedback based retrieval (see [12] and references therein). There, multiple positive examples are selected by the user on-line to represent the specific class of images being retrieved; negative (irrelevant) examples can also be selected, but used as class non-specific. Therefore, the way of doing classification in relevance feedback is "one-versus-the-rest". In contrast, the NFL method presented in this paper can be considered as taking just one positive example, *i.e.* the query, with no negatives and with no feedback.

B. The Feature Line Space

In the NFL method, a feature subspace is constructed for each class, consisting of the straight lines (feature lines) passing through each pair of the prototypes (feature points) belonging to that class. The prototypes are generalized by the feature lines.

Consider a variation in the image space from point $\mathbf{z}_1$ to $\mathbf{z}_2$ and the corresponding variation in the feature space, which in this work is that of Gabor and wavelet features, from $\mathbf{x}_1$ to $\mathbf{x}_2$. The degree of the change may be measured by $\delta\mathbf{z} = \|\mathbf{z}_2 - \mathbf{z}_1\|$ or $\delta\mathbf{x} = \|\mathbf{x}_2 - \mathbf{x}_1\|$. When $\delta\mathbf{z} \to \mathbf{0}$ and thus $\delta\mathbf{x} \to \mathbf{0}$, the locus of $\mathbf{x}$ due to the change can be approximated well enough by a straight line segment between $\mathbf{x}_1$ and $\mathbf{x}_2$. Thus any variant between the two can be interpolated by a point on the line. A further small change beyond $\mathbf{x}_2$ can be extrapolated using the linear model.

The straight line passing through $\mathbf{x}_1$ and $\mathbf{x}_2$ of the same class, denoted by $\overline{\mathbf{x}_1\mathbf{x}_2}$, is called a *feature line* (FL) of that class (see Fig.1). The FL provides information about linear variants of the two prototypes, *i.e.* possible images derived from the two. It virtually provides an infinite number of prototype feature points of the class that the two prototypes belong to. The prototypical set of a class is thus expanded by the FL subspace.

Let $\mathbf{x}^c = \{\mathbf{x}_i^c \mid 1 \leq i \leq N_c\}$ be the set of N_c prototypical feature points belonging to class c. A number of $K_c = \frac{N_c(N_c-1)}{2}$ FLs can be constructed to represent the class. For example, $N_c = 5$ feature points are expanded by their $K_c = 10$ FLs. All the K_c FLs constitute the FL space of class c, $\mathbf{S}^c = \{\overline{\mathbf{x}_i^c \mathbf{x}_j^c} \mid 1 \leq i, j \leq N_c, i \neq j\}$, which is a subset of the entire feature space. When there are M classes in the database, M such FL spaces can be constructed, composed of a total number of $N_{total} = \sum_{c=1}^{M} K_c$ FL's.

A FL covers more of the feature space than the two feature points along, but this expansion of feature set is constrained by the original feature points. It virtually provides an infinite number of feature points derived from the original feature points, accounting for more yet constrained possibilities than the original points. The generalization ability of the classifier is thus increased. The distance between the query vector and its projection onto the subspace is calculated and used as the metrics.

The loci of the feature points of an image under perceivable variations in viewpoint, illumination or expression are highly nonconvex and complex [13]. To obtain a fine description of the variations, one may suggest that a higher order curve, such as splines, should be used, as did in [11] in strictly controlled situations where the images can be ordered in terms of a single parameter (such as rotation angle on a turnable). This requires (i) that there should be at least three prototypical points for every class, and (ii) that these points should be ordered to account for relative variations described by only one parameter. In image classification and retrieval, requirement (ii) is generally not satisfied; this is because the diversity among the prototype images is great, much more than variations in viewpoint, illumination and so on. Moreover, an ordering of images is impossible for the task here because requirement (ii) cannot be satisfied.

The NFL method generalizes individual prototypes $\mathbf{x}^c$ by constructing a simplified manifold, that is, the FL space. Although the FL space is a crude approximation for representing variations within an image class, it turns out to be useful for the classification and retrieval when used with the NFL criterion to be described in the following, and can achieve significant improvements over conventional methods such as the NN. The NFL presents an effective mean of using the constraint from multiple prototypes per class, as can be seen in comparison with k-NN and nearest center methods.

C. Image Classification/Retrieval Using NFL

For the *NFL classification*, a query feature point $\mathbf{x}$ is classified to c if it is nearest to $\mathbf{S}^c$ (The distance from $\mathbf{x}$ to $\mathbf{S}^c$ is the shortest distance from $\mathbf{x}$ to the FL's belonging to $\mathbf{S}^c$). For the *NFL retrieval*, two patterns represented by $\mathbf{x}_i^c$ and $\mathbf{x}_j^c$ are retrieved as the top two if $\mathbf{x}$ is closest to $\overline{\mathbf{x}_i^c\mathbf{x}_j^c}$; other pairs can be retrieved and ranked according to the FL distance defined below.

Letting $\mathbf{p}$ be the projection point of the query $\mathbf{x}$ onto $\overline{\mathbf{x}_1\mathbf{x}_2}$ (see Fig.1), the *FL distance* from $\mathbf{x}$ to $\overline{\mathbf{x}_1\mathbf{x}_2}$ is defined as $d(\mathbf{x}, \overline{\mathbf{x}_1\mathbf{x}_2}) = \|\mathbf{x} - \mathbf{p}\|$ where $\|\cdot\|$ is some norm. The projection point can be computed as $\mathbf{p} = \mathbf{x}_1 + \mu(\mathbf{x}_2 - \mathbf{x}_1)$ where $\mu \in \mathcal{R}$, called the position parameter, can be calculated from $\mathbf{x}$, $\mathbf{x}_1$ and $\mathbf{x}_2$ as follows: Because $\overline{\mathbf{px}}$ is perpendicular to $\overline{\mathbf{x}_2\mathbf{x}_1}$, we have $(\mathbf{p} - \mathbf{x}) \cdot (\mathbf{x}_2 - \mathbf{x}_1) = [\mathbf{x}_1 + \mu(\mathbf{x}_2 - \mathbf{x}_1) - \mathbf{x}] \cdot (\mathbf{x}_2 - \mathbf{x}_1) = 0$ where "$\cdot$" stands for dot product, and thus $\mu = \frac{(\mathbf{x}-\mathbf{x}_1)\cdot(\mathbf{x}_2-\mathbf{x}_1)}{(\mathbf{x}_2-\mathbf{x}_1)\cdot(\mathbf{x}_2-\mathbf{x}_1)}$. The parameter μ describes the position of $\mathbf{p}$ relative to $\mathbf{x}_1$ and $\mathbf{x}_2$. When $\mu = 0$, $\mathbf{p} = \mathbf{x}_1$. When $\mu = 1$, $\mathbf{p} = \mathbf{x}_2$. When $0 < \mu < 1$, $\mathbf{p}$ is an interpolating point between $\mathbf{x}_1$ and $\mathbf{x}_2$. When $\mu > 1$, $\mathbf{p}$ is a "forward" extrapolating point on the $\mathbf{x}_2$ side. When $\mu < 0$, $\mathbf{p}$ is a "backward" extrapolating point on the $\mathbf{x}_1$ side.

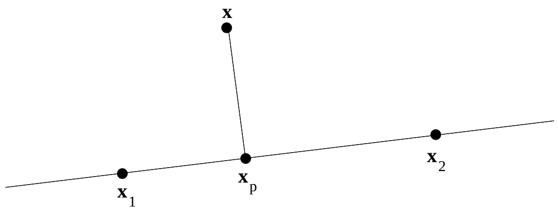


Fig. 1. Generalizing two feature points $\mathbf{x}_1$ and $\mathbf{x}_2$ by the feature line $\overline{\mathbf{x}_1\mathbf{x}_2}$. The feature point $\mathbf{x}$ of a query is projected onto the line as point $\mathbf{p}$.

The NFL computation procedure follows: Calculate the FL distance between the query $\mathbf{x}$ and the feature line $\overline{\mathbf{x}_i^c\mathbf{x}_j^c}$ for each class c and each pair (i,j) where $i \neq j$. This yields a number of N_{total} distances. The distances are sorted in ascending order, each being associated with a class identifier c, two prototypes $\mathbf{x}_i^c$ and $\mathbf{x}_j^c$, and the corresponding μ value. The *NFL distance* is the first rank FL distance: $d(\mathbf{x}, \overline{\mathbf{x}_{i^*}^{c^*}\mathbf{x}_{j^*}^{c^*}}) = \min_{1\le c\le M} \min_{1\le i<j\le N_c} d(\mathbf{x}, \overline{\mathbf{x}_i^c\mathbf{x}_j^c})$. The first rank gives the information about the best matched class c^* for the NFL classification, and the two associated prototypes i^* and j^* for the NFL retrieval.

III. Performance Evaluation

Experiments were conducted to evaluate the performance of the NFL against three other existing image classification/retrieval methods, namely, the NN, the nearest class center (NC), and the k-NN [1]. The NC and the k-NN also make use of class information. Gabor and wavelet feature representations are adopted and used as the common starting point for the comparison among several search methods. Comparisons are made among various combinations of methods and feature representations. For example, the NFL with Gabor feature is denoted as NFL+Gabor, and so on.

The performance of all the compared methods will be evaluated in terms of the following two measures:

1. *Error rate.* This is a measure for classification calculated as the number of incorrect first rank matches divided by the total number of queries.

2. *Retrieval efficiency.* This was proposed in [14] and also adopted *e.g.* in [4]. It is defined as $\eta_w(q,m) = \sum_{k=1}^{m} \frac{1}{N_q} Match(q, r_k)$ where q is the query and $r_1, \ldots, r_m$ are the m top ranked matches for the query q; $Match(q, r_k) = 1$ if r_k and q belong to the same class, or 0 otherwise; and N_q is the number of available prototypes for the class that q belongs to. The highest possible efficiency value is 1 when all the top N_q matches are correct. The average retrieval efficiency over all q is finally used.

When an image is used as the query, it is *not* used as a prototype, *i.e.* it is removed from the prototype set, during the classification (the leave-one-out test).

A. Brodatz Textures

The Brodatz texture database consisting of 112 different monochrome texture images of size 512×512 is used for the evaluation. This set of experiments follow a protocol similar to that devised in [4]: Each image is divided into 16 non-overlapping subimages of size 128×128. This creates a total number of $112 \times 16 = 1792$ subimages in the database. Each subimage is then used as the query q, with the other $N_c = N_q = 15$ subimages as the multiple prototypes.

Gabor and wavelet features are extracted from the images. Gabor features are extracted using Gabor Filters at 4 scales and 6 orientations. The Gabor feature set used in these tests is exactly the same as that used by [4], which is downloaded from `http://vivaldi.ece.ucsb.edu/users/wei/gaborfeatures` (This data set contains 4 other textures in addition to the 112 classes). The mean and the standard deviation of the magnitude of the filtered images are calculated as features representing a subimage. Wavelet features are extracted from the *Lab* space by a 3-level wavelet decomposition using a mother wavelet (the "Daub4" wavelet). The mean and the standard deviation of the approximate and detailed coefficients are calculated as features representing a subimage. Hence there are a total of 48 Gabor features or 24 wavelet features for an image. The mean and deviation components are then normalized by the standard deviations of the respective components over the entire database.

The classification error rates with the wavelet features are compared in Table I, which shows that the NFL achieves the lowest error rates than the other methods. Fig.2 shows the retrieval efficiencies. The NFL produces consistently better results than the NC and NN in the retrieval efficiencies for every m values and for both Gabor and wavelet feature representations (the best retrieval efficiency reported in [4] was obtained by using the NN+Gabor method). The NFL achieves 90.00% retrieval efficiency as opposed to 74.37% of the NN search as reported in [4] when top 15 matches are considered.

B. General Color Images

The NFL method is also applied to the classification and retrieval of color images, using a database consisting of 1264 color images from the MIT VisTex database and some images from Corel Stock Photos. The images are classified into 71 classes manually and subjectively by human observers.

All these color images are resized to 128 by 128 pixels and then processed by firstly

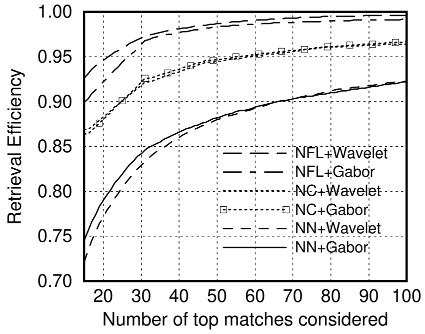
TABLE I

ERROR RATES (IN %) FOR BRODATZ TEXTURE (T) AND COLOR IMAGE (C) IMAGES USING GABOR (G) AND WAVELET (W) FEATURES.

Method	NN	5-NN	10-NN	15-NN	NC	NFL
Error rate (TW)	6.2	8.6	10.9	14.6	13.1	5.3
Error rate (TG)	10.8	10.9	13.4	15.5	13.6	7.9
Error rate (CW)	45.0	53.4	53.8	57.1	60.4	42.0
Error rate (CG)	49.6	56.1	59.8	60.3	62.1	44.4

converting their RGB values to the perceptual uniform $CIE - Lab$ color space. Gabor and wavelet features are extracted (in the same way as in the monochrome case) from the L, a and b images, separately. The mean and standard deviation values are calculated for each color attribute image. The three feature vectors are concatenated, forming a feature vector of 48×3 dimensions for Gabor features or 24×3 for wavelet features. The feature components are normalized as before.

Table I and Fig.2 show the results. Given the same set of features, the NFL still produces



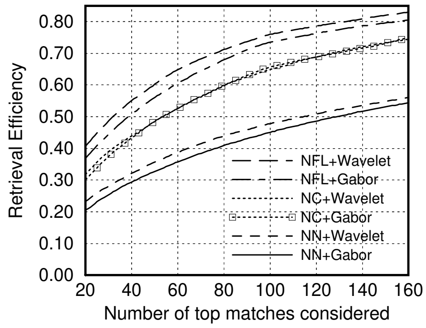


Fig. 2. Retrieval efficiencies for Brodatz textures of size 128x128 (left) and for color images (right).

significantly better results than other compared methods: The NFL has a 75% of increase in retrieval efficiency over the NN method when top 20 matches are considered. When top 100 matches are considered, the retrieval efficiencies are 0.760 for NFL+Wavelet, 0.735 for NFL+Gabor, 0.478 for NN+Wavelet, and 0.451 for NN+Gabor.

Note that for the above two set of experiments with the Brodatz and color image databases, the other two methods, NC and k-NN, which also make use of multiple prototype information, are even inferior to the simple NN in terms of the error rate. However, the NC is better than the NN in retrieval efficiency where a larger number of top matches are considered. This demonstrates that the NFL makes best use of information about multiple prototypes.

In this set of experiments, however, the error rates are high and the retrieval efficiencies are low for all the methods. This is because the low level features do not necessarily correlate to the human perception of the color image content. A clustering analysis in the Gabor and wavelet feature spaces indicates that the feature clusters are not well consistent with the image content classification. To achieve better content-based classification and retrieval, a way of bridging the gap between low level features and high level subjective perception has to be devised. This issue are being addressed. Strategies such as relevance feedback [12] may help in this regard.

C. With Size Reduction and Sub-Sampling

Now we evaluate how the compared methods behave when the image size is reduced and sub-sampled in a similar way as [15]. First, the previous 112 texture images of size 128x128 are reduced to the sizes of 64x64, 32x32 and 16x16, respectively. Figs 3 shows

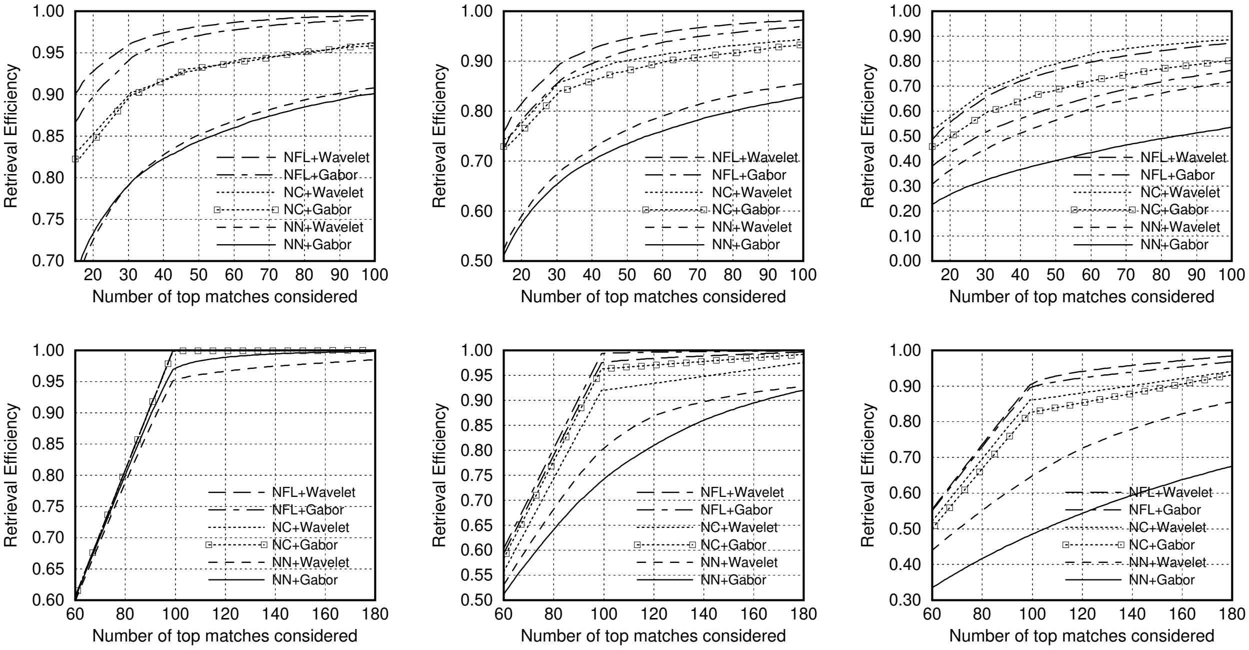


Fig. 3. Retrieval efficiencies with reduced size (top) and sub-sampled (bottom) images of 64x64, 32x32 and 16x16 pixels (from left to right).

the retrieval efficiencies for the three reduced sizes. When the size is reduced to 16x16, the NC performs slightly better (by about 3%) than the NFL; but this does not indicate that the NC is a better method because in this case the appearances of texture details of most images are actually lost.

Second, the experiments are done with sub-sampled data for the nine classes from Brodatz's album as chosen by [15]. The original texture images are randomly sub-sampled at 64x64, 32x32 and 16x16 pixels. 100 subsamples are extracted from each size and each class, resulting in a database of 900 random samples in total at each image size. The results are shown in Fig.3. The performances of all methods drop as the image resolution is reduced. This is due to the loss of image details. However, the NFL still performs the best of all.

IV. Conclusion

A novel distance metric, the NFL distance, is proposed for image classification and retrieval. The NFL makes use of available information about multiple prototypes within a class by constructing a subspace that describe the variations of features within a class. The experimental results have shown that given the same set of features, the NFL consistently achieves better performance than the NN and as well as the other two methods, k-NN and NC, which also use the multiple prototype information, whether the Gabor or wavelet features are used. This demonstrates that when multiple prototypes are available, the NFL makes efficient use of the class information.

The NFL turns out to be a general pattern recognition method, regardless of representations, and is applicable when there are at least two prototypes per class. Our recent research shows that the NFL outperforms the compared methods also in other applications such as in face recognition [2], and in audio classification and retrieval [16] (see demos at `http://research.microsoft.com/~szli/demos.html`).

The NFL must have assumed (implicitly) some forms of correlations between prototypes within a class. However, the forms which the NFL takes advantage of (whereas NN, k-NN and NC do not) may not be easily identified. We are still investigating why and how and developing a theory to justify the NFL concept.

Acknowledgment

We would like to thank Dr. W.Y. Ma's for his suggestions for reconstructing the Gabor filters used in [4] for the performance comparison.

References

[1] K. Fukunaga, *Introduction to statistical pattern recognition*, Academic Press, Boston, 2 edition, 1990.

[2] S. Z. Li and J. Lu, "Face recognition using the nearest feature line method", *IEEE Transactions on Neural Networks*, vol. 10, no. 2, pp. 439–443, March 1999.

[3] T. Chang and C.-C. Kuo, "Texture analysis and classification with tree-structured wavelet transform", *IEEE Transactions on Image Processing*, vol. 3, no. 4, pp. 329–339, 1993.

[4] B. S. Manjunath and W. Y. Ma, Texture features for browsing and retrieval of image data, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, August 1996.

[5] M. K. Mandal, T. Aboulnasr, and S. Panchanathan, Image indexing using moments and wavelets, *IEEE Transactions on Consumer Electronics*, vol. 42, no. 3, pp. 557–565, 1996.

[6] P. Brodatz, *"Textures: A photographic album for artists and designers"*, Dover Publications, Inc., New York, 1966.

[7] S. Ullman and R. Basri, "Recognition by linear combinations of models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 992–1006, 1991.

[8] T. Poggio and S. Edelman, "A network that learn to recognize three-dimensional objects", *Nature*, vol. 343, pp. 263–266, 1990.

[9] Shimon Edelman and Sharon Duvdevani-Bar, "A model of visual recognition and categorization", *Proceedings of Royal Society, London*, vol. B-352, pp. 1191–1202, 1997.

[10] Thomas Vetter and Tomaso Poggio, "Linear object classes and image synthesis from a single example image", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 733–742, 1997.

[11] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance", *International Journal of Computer Vision*, vol. 14, pp. 5–24, 1995.

[12] Yong Rui, Thomas S. Huang, Michael Ortega, and Sharad Mehrotra, "Relevance feedback: A power tool for interactive content-based image retrieval", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, pp. 644–655, 1998.

[13] M. Bichsel and A. P. Pentland, "Human face recognition and the face image set's topology", *CVGIP: Image Understanding*, vol. 59, pp. 254–261, 1994.

[14] B. M. Mehtre, M. S. Kankanhalli, A. D. Narasimhalu, and G. C. Man, "Color matching for image retrieval", *Pattern Recognition Letters*, vol. 16, pp. 325–331, 1995.

[15] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions", *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, January 1996.

[16] S. Z. Li, "Content-based classification and retrieval of audio using the nearest feature line method", *IEEE Transactions on Speech and Audio Processing*, September 2000.