# NOVEL GENERAL KNN CLASSIFIER AND GENERAL NEAREST MEAN CLASSIFIER FOR VISUAL CLASSIFICATION

*Qingfeng Liu, Ajit Puthenputhussery, Chengjun Liu*

Department of Computer Science
New Jersey Institute of Technology

## ABSTRACT

This paper presents a novel general k nearest neighbour classifier (GKNNc) and a novel general nearest mean classifier (GNMc) for visual classification. Instead of treating the data equally, both GKNNc and GNMc assign a weight coefficient to each data. To achieve good performance, the conditions and properties of the weight coefficients for GKNNc and GNMc are further analysed. Then a sparse representation based method is proposed to derive the weight coefficients for both GKNNc and GNMc. Experimental results on several representative data sets, such as the Caltech 101 dataset and the MIT-67 indoor scenes dataset demonstrate the feasibility of the proposed methods.

## 1. INTRODUCTION

The k nearest neighbour classifier (KNNc) and the nearest mean classifier (NMc) are widely used for visual classification [1]. The main problem of KNNc and NMc is that they treat all the data samples equally, which is error-prone in practice. Some earlier generalization work [2], [3] on KNNc have been done by using the theory of fuzzy sets. However, the robustness is not fully considered and generalization on NMc is not developed. Sparse representation [4], [5] methods have shown its robustness in several visual classification cases, e.g. face recognition, object recognition.

In this paper, a novel general k nearest neighbour classifier (GKNNc) and a novel general nearest mean classifier (GNMc) are presented from a novel sparse representation point of view rather than the fuzzy sets point of view [2], [3].

Both GKNNc and GNMc admit the importance of the choice of the weight coefficients for the training samples in order to achieve good performance. Specifically, two conditions, namely the Bayes decision rule condition and the robustness condition, are proposed for the weight coefficients for both GKNNc and GNMc . The Bayes decision rule condition aims at establishing a connection between GKNNc and the Bayes decision rule for minimum error. To achieve so, the closer training samples are assigned larger weight coefficient by the means of kernel density estimation. Such a condition will also benefit GNMc since the new definition of "mean" for each class concentrates more on the important training samples for GNMc. The robustness condition applies the $L_1$ norm and discards the distant noisy samples so that the weight coefficient vector is sparse. As per these two conditions, a sparse representation based method (SRBM) is further proposed to derive the satisfying weight coefficients for both GKNNc and GNMc.

The proposed GKNNc and GNMc are then evaluated on four representative databases: the extended Yale face database B [6], the 15 scenes dataset [7], the MIT-67 indoor scenes dataset [8] and the Caltech 101 dataset [9]. Experimental results show the competence of the proposed classifiers compared to the linear or non-linear support vector machine.

The key contributions of this paper are summarized as follows: (i) a novel GKNNc and a novel GNMc are proposed; (ii) two conditions on the weight coefficients are proposed for better performance. and (iii) a novel sparse representation based method is proposed to derive the weight coefficients by combining both GKNNc and GNMc into the objective function.

## 2. RELATED WORK

Some earlier generalization work [2], [3] on KNNc and NMc have been done. Keller et al. proposed a fuzzy k-nearest neighbour algorithm by incorporating the theory of fuzzy sets to assign the fuzzy memberships to the training samples. Bezdek et al. extended the idea of fuzzy KNN to a general framework and conducted both theoretical and empirical analysis.

Sparse representation method [4] was proposed for robust face recognition and further applied to other recognition tasks such as object recognition, scene recognition, and action recognition. To further improve performance, some discriminative dictionary learning methods have been proposed for sparse representation. Zhang et al. [10] proposed a similar objective function and applied a discriminative singular value decomposition (D-KSVD) method to learn the discriminative dictionary and the classifier simultaneously. Jiang et al. [11] improved upon the method introduced in [10] by introducing a label consistent regularization term. Yang et al. [12], [13] proposed a Fisher Discrimination Dictionary

Learning (FDDL) method, which learns a structured dictionary that consists of a set of class-specific sub-dictionaries.

## 3. GKNNC AND GNMC

This section presents a general k nearest neighbour classifier (GKNNc) and a general nearest mean classifier (GNMc).

The GKNNc is defined as follows given the training sample matrix $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_m] \in \mathbb{R}^{n \times m}$ and the test sample $\mathbf{x} \in \mathbb{R}^n$,

$$c^* = \arg\max_c \sum_{\mathbf{t}_i \in \mathbf{T}_c} w_i \quad (1)$$

where $c = 1, 2, ..., l$ is the class label, $w_i$ is the corresponding weight coefficient of training sample $\mathbf{t}_i$, $\mathbf{T}_c$ is the set of training samples in the $c$-th class and $\sum_{\mathbf{t}_i \in \mathbf{T}_c} w_i$ is the sum of the weight coefficients of the training samples in the $c$-th class. It is obvious that the GKNNc becomes the traditional KNN classifier if $w_i = 1$ when $\mathbf{t}_i$ is among the nearest neighbours of $\mathbf{x}$ and $w_i = 0$ otherwise.

Correspondingly, the GNMc is defined as follows given the same notations,

$$c^* = \arg\min_c ||\mathbf{x} - \sum_{\mathbf{t}_i \in \mathbf{T}_c} w_i \mathbf{t}_i||_2^2 \quad (2)$$

where $\sum_{\mathbf{t}_i \in \mathbf{T}_c} w_i \mathbf{t}_i$ is the "mean" of the $c$-th class. It can be found that the GNMc becomes the traditional nearest mean classifier if $w_i$ is the inverse of the size of $\mathbf{T}_c$ for all the training samples in $\mathbf{T}_c$.

For example, as shown in Fig.1 there are four training samples $\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3$ and $\mathbf{t}_4$ with class labels 1, 1, 2, 2. The representation $\mathbf{w} = [0.6, 0.6, 0.0, 0.8]$ is derived by means of some methods for the test sample $\mathbf{x}$. Then the GKNNc will classify the test sample to class 1 since $0.6 + 0.6 > 0.0 + 0.8$. While the GNMc will compute $r_1 = ||\mathbf{x} - (0.6\mathbf{t}_1 + 0.6\mathbf{t}_2)||^2$ and $r_2 = ||\mathbf{x} - (0.0\mathbf{t}_3 + 0.8\mathbf{t}_4)||^2$ and choose the smaller one.

It can be revealed that the weight coefficient for each training sample plays an important role in the classification performance. Therefore, some conditions for the weight coefficients are proposed to guarantee the classification performance. And based on such conditions, a sparse representation based method is further proposed to derive the weight coefficients.

### 3.1. Conditions

Since the weight coefficient plays an important role in the classification performance, it is necessary to impose the following conditions on the weight coefficients.

- **The Bayes decision rule condition**. If the test sample $\mathbf{x}$ is to the training sample $\mathbf{t}_i$, the larger $w_i$ is. Particularly, the following method is applied to compute the distance between the test sample $\mathbf{x}$ and the training sample $\mathbf{t}_i$

$$d_i = exp\{-\frac{1}{2\sigma^2}||\mathbf{x} - \mathbf{t}_i||^2\} \quad (3)$$

where the parameter $\sigma$ is used to adjust the decay speed. Then the distance vector $\mathbf{d} = [d_1, d_2, ..., d_m]^t \in \mathbb{R}^m$ can be constructed.

Ideally, the weight coefficient $w_i$ has a proportional relation to $d_i$ so that $w_i \approx \beta d_i + constant$ because the following favourable property for GKNNc is obtained in such case from the kernel density estimation point of view,

$$\begin{aligned} c^* &= \arg\max_c \sum_{\mathbf{t}_i \in \mathbf{T}_c} w_i \\ &\approx \arg\max_c p(\mathbf{x}|c) \\ &\propto \arg\max_c p(c|\mathbf{x}) \quad \text{if (p(c) is equal)} \end{aligned} \quad (4)$$

It is found that GKNNc has a connection to the Bayes decision rule for minimum error under this condition. Moreover, the "mean" defined in GNMc is more reasonable under this condition.

- **The robustness condition**. The weight coefficient vector $\mathbf{w} = [w_1, w_2, ..., w_m]^t \in \mathbb{R}^m$ should be sparse by applying $L_1$ norm and discarding distant neighbors. The reason is that $L_1$ norm is robust and not all training samples $\mathbf{t}_i$ are required for classification since most of them are noisy, especially the distant samples. Under this condition the robustness of the proposed GKNNc and GNMc is improved.

### 3.2. Learning the Weight Coefficients

According to the conditions and properties, a sparse representation based method (SRBM) for deriving the weight coefficients is presented in this section. The following objective function is proposed by combining both GKNNc and GNMc together,

$$\min_{\mathbf{w}} ||\mathbf{x} - \mathbf{T}\mathbf{w}||^2 + \lambda||\mathbf{w}||_1 + \alpha||\mathbf{w} - \beta\mathbf{d}||^2 \quad (5)$$

where the parameter $\lambda$ controls the sparseness, $\alpha$ trades off between GKNNc and GNMc, $\beta$ constrains the value of vector $\mathbf{d}$, $||\cdot||$ is the $L_2$ norm and $||\cdot||_1$ is the $L_1$ norm. The SRBM is capable of deriving the weight coefficient vector $\mathbf{w}$ that satisfies both of the two conditions.

The FISTA (Fast Iterative Shrinkage Thresholding Algorithm) algorithm [14] is applied for optimizing the criterion in equation 5. The equation 5 is decomposed into the following form: $f(\mathbf{w}) + g(\mathbf{w})$, where $f(\mathbf{w}) = ||\mathbf{x} - \mathbf{T}\mathbf{w}||^2 + \alpha||\mathbf{w} - \beta\mathbf{d}||^2$ and $g(\mathbf{w}) = \lambda||\mathbf{w}||_1$ to meet the condition of FISTA algorithm. To guarantee convergence, the maximal

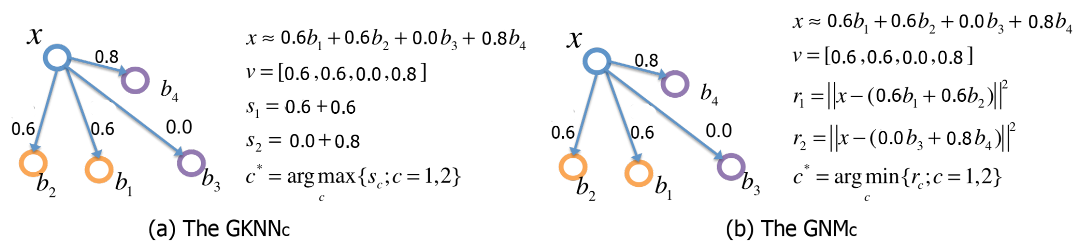


**Fig. 1**. Illustration example of two proposed classifiers: the GKNNc and the GNMc

step size for the FISTA algorithm is selected as $\frac{1}{L}$, where $L = 2\lambda_{max}(\mathbf{T}^t\mathbf{T} + \alpha\mathbf{I})$, which means twice the largest eigenvalue of the matrix $\mathbf{T}^t\mathbf{T} + \alpha\mathbf{I}$.

## 4. EXPERIMENTS

In this section, the performance of our proposed classifiers GKNNc and GNMc is assessed on several visual classification databases: the extended Yale face database B [6], the 15 scenes dataset [7], the MIT-67 indoor scenes dataset [8] and the Caltech 101 dataset [9]. The image or video is first represented as a pattern vector. As for the extended Yale face database B, the pattern vector is first formed as the concatenation of the column pixels. Then the random faces [4], which is the row vectors of a randomly generated transformation matrix from a zero-mean normal distribution, is applied to project the face pattern vector into a dimension of 540 representation vector. Each row of the transformation matrix is normalized to unit length. For the 15 scenes dataset, the spatial pyramid feature provided by [11] is used to represent the image for fair comparison. The feature is obtained by using a four-level spatial pyramid and a codebook with a size of 200. For the MIT-67 indoor scenes dataset, the Fisher vector feature [15] is used to represent the image. For the Caltech 101 dataset, the proposed method is built upon the 4096 dimension image features that are extracted by using a pre-trained convolutional neural network CNN-M [16]. Please see more details about the features in the corresponding sub-sections. Then the marginal Fisher analysis preceded with the principal component analysis (PCA) [17] is applied to reduce the dimension and extract features. Afterwards, both GKNNc and GNMc are applied for classification according to the experimental settings defined for different datasets.

### 4.1. Extended Yale Face Database B

The extended Yale face database B consists of 2414 frontal view face images from 38 individuals each with around 64 images taken under various lightening conditions. A cropped version of the database is used wherein all the images are manually aligned, cropped, and then re-sized to $168 \times 192$ [6]. Following the experimental setting [12], 20 images are randomly selected for training for each subject, and the remaining images for testing for 10 iterations. Note that this experimental setting is more difficult than that in [10]. The image is first scaled to $42 \times 48$. The random faces [4] is used to obtain the pattern vector to prove the robustness of the proposed method. Then the dimension is reduced to 350. The parameters are selected as $\sigma = 1$, $\lambda = 0.02$, $\alpha = 0.1$, and $\beta = 0.5$. The final results are shown in table 2. Our proposed method significantly improves upon the other popular methods by more than 4 percent.

### 4.2. The 15 Scenes Dataset

The 15 scenes dataset [7] contains totally 4485 images from 15 scene categories, each with the number of images ranging from 200 to 400. Following the experimental protocol defined in [7], 100 images per class are randomly selected for training and the remaining for testing for 10 iterations. First, the spatial pyramid features provided by [11] is used to represent the image as a vector with a dimension of 3000 for fair comparison. Then the dimension is further reduced to 500. The parameters are selected as $\lambda = 0.05$, $\alpha = 0.1$, and $\beta = 1.0$. It can be concluded from the results in table 1 that our proposed method is able to achieve much better results than the non-linear or linear kernel based support vector machine that is used by the compared methods.

### 4.3. The MIT-67 Indoor Scenes Dataset

The MIT-67 indoor scenes dataset [8] is a very challenging indoor scene recognition dataset, which contains 67 indoor categories with 15620 images. According to the commonly used experimental setting [8], 80*67 images are used for training and 20*67 images for testing. The performance is measured as the average classification accuracy over all the categories. The Fisher vector feature [15] is considered for representing the image. The SIFT feature is first projected to 80 dimension and a codebook with 256 visual words is computed, then the dimension of the Fisher vector is 2*256*80 = 40960, which is further reduced to 2000. The parameters are selected as $\lambda = 0.01$, $\alpha = 0.1$, and $\beta = 1.5$ for the GNMc while $\beta = 0.5$ for the GKNNc to achieve the best performance.

| Methods | Accuracy % |
|---|---|
| LLC [19] | 80.57 |
| D-KSVD [10] | 89.10 |
| LC-KSVD [11] | 90.40 |
| LaplacianSC [20] | 89.7 |
| **The proposed GNMc** | **97.45**$\pm$0.27 |
| **The proposed GKNNc** | **93.54**$\pm$0.45 |

**Table 1**. Comparisons between the proposed GKNNc, GNMc and the other popular methods on the 15 scenes dataset

| Methods | Accuracy % |
|---|---|
| D-KSVD [12] | 75.30 |
| SRC [12] | 90.00 |
| FDDL [12] | 91.90 |
| **The proposed GNMc** | **95.35** |
| **The proposed GKNNc** | **95.39** |

**Table 2**. Comparisons between the proposed GKNNc, GNMc and the other popular methods on the extended Yale face database B.

The results are shown in table 3. Please note that the Fisher vector is learned directly from the SIFT features of images instead of learning the part detectors in [18], which is time-consuming. Moreover, the dimension of Fisher vector is reduced from 40960 to 2000, which saves much storage space and no data augmentation technique is used. However, the proposed methods can still achieve very competitive results to the state-of-the-art methods [18] that use support vector machine.

### 4.4. The Caltech 101 Dataset

The Caltech 101 dataset [26] holds 9144 images divided into 101 object classes and a clutter class. To be consistent with the previous work [19], we partition the whole dataset randomly into 5, 10, 15, 20, 25, 30 training images per class and no more than 50 test images per class, and measure the performance using the average accuracy over 102 classes. In order to achieve comparable results to the state-of-the-art methods, the proposed method is built upon the 4096 dimension image representation features that are extracted by using a pre-trained convolutional neural network CNN-M [16]. Then we reduce the dimension to 1000 except the case with 5 training images, where the dimension is reduced to 500. The model parameters are selected as $\sigma = 1.5$, $\lambda = 0.05$, $\alpha = 0.1$, and $\beta = 1.5$. For the shifted power transformation, $\lambda_1 = 0.0$ and $\lambda_2 = 0.5$. For the GKNNc, the value of $k = 20$ if the training image size is larger than 20, otherwise k is the value of the training image size. The same applies to the GNMc. The results that are shown in table 4, demonstrate that the proposed method is comparable to the state-of-the-art methods with support vector machine for classification in different training image sizes.

| Methods | Mean Accuracy % |
|---|---|
| ROI + Gist [8] | 26.1 |
| DPM [21] | 30.4 |
| Object Bank [22] | 37.6 |
| miSVM [23] | 46.4 |
| D-Parts [24] | 51.4 |
| DP + IFV [18] | 60.8 |
| CNN-SVM no Aug [25] | 58.4 |
| **The proposed GNMc** | **59.12** |
| **The proposed GKNNc** | **58.40** |

**Table 3**. Comparisons between the proposed GKNNc, GNMc and the other popular methods on the MIT-67 indoor scenes dataset.

| training images | 15 | 20 | 25 | 30 |
|---|---|---|---|---|
| LLC [19] | 65.43 | 67.74 | 70.16 | 73.44 |
| SRC [4] | 64.90 | 67.70 | 69.20 | 70.70 |
| D-KSVD [10] | 65.10 | 68.60 | 71.10 | 73.00 |
| LC-KSVD [11] | 67.70 | 70.50 | 72.30 | 73.60 |
| Zeiler [27] | 83.80 | - | - | 86.5 |
| CNN-M + Aug [16] | - | - | - | 87.15 |
| **The proposed GNMc** | **84.79** | **85.96** | **86.62** | **87.68** |
| **The proposed GKNNc** | **84.76** | **86.11** | **86.77** | **87.74** |

**Table 4**. Comparisons between the proposed method and the other popular methods on the Caltech 101 dataset.

### 4.5. Conclusion

This paper presents two novel classifiers: GKNNc and GNMc for visual classification. Two conditions of the weight coefficients for GKNNc and GNMc are analysed to guarantee the performance. Then a sparse representation based method is proposed to derive the weight coefficients for both GKNNc and GNMc. Experimental results on several representative data sets demonstrate the feasibility of the proposed methods.

## 5. REFERENCES

[1] Keinosuke Fukunaga, *Introduction to Statistical Pattern Recognition (2Nd Ed.)*, Academic Press Professional, Inc., 1990.

[2] J C Bezdek, S K Chuah, and D Leep, "Generalized k-nearest neighbor rules," *Fuzzy Sets Syst.*, vol. 18, no. 3, pp. 237–256, 1986.

[3] James M. Keller, Michael R. Gray, and James A. Givens, "A fuzzy k-nearest neighbor algorithm.," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, no. 4, pp. 580–585, 1985.

[4] John Wright, Allen Y. Yang, Arvind Ganesh, Shankar S. Sastry, and Yi Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009.

[5] Bin Cheng, Jianchao Yang, Shuicheng Yan, Yun Fu, and Thomas S. Huang, "Learning with l1-graph for image analysis," *Trans. Img. Proc.*, vol. 19, no. 4, pp. 858–866, 2010.

[6] K.C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. 27, no. 5, pp. 684–698, 2005.

[7] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *CVPR*, 2006, pp. 2169–2178.

[8] Ariadna Quattoni and Antonio Torralba, "Recognizing indoor scenes," in *CVPR*, 2009, pp. 413–420.

[9] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," Tech. Rep. 7694, California Institute of Technology, 2007.

[10] Qiang Zhang and Baoxin Li, "Discriminative k-svd for dictionary learning in face recognition.," in *CVPR*, 2010, pp. 2691–2698.

[11] Zhuolin Jiang, Zhe Lin, and Larry S. Davis, "Label consistent k-svd: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, 2013.

[12] Meng Yang, Lei Zhang, Xiangchu Feng, and David Zhang, "Fisher discrimination dictionary learning for sparse representation," in *ICCV*, 2011, pp. 543–550.

[13] Meng Yang, Lei Zhang, Xiangchu Feng, and David Zhang, "Sparse representation based fisher discrimination dictionary learning for image classification," *International Journal of Computer Vision*, pp. 1–24, 2014.

[14] Amir Beck and Marc Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[15] Jorge Sanchez, Florent Perronnin, Thomas Mensink, and Jakob J. Verbeek, "Image classification with the fisher vector: Theory and practice.," *International Journal of Computer Vision*, vol. 105, no. 3, pp. 222–245, 2013.

[16] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in *BMVC*, 2014.

[17] Shuicheng Yan, Dong Xu, Benyu Zhang, HongJiang Zhang, Qiang Yang, and Stephen Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, 2007.

[18] Mayank Juneja, Andrea Vedaldi, C. V. Jawahar, and Andrew Zisserman, "Blocks that shout: Distinctive parts for scene classification," in *CVPR*, 2013, pp. 923–930.

[19] Jinjun Wang, Jianchao Yang, Kai Yu, Fengjun Lv, Thomas S. Huang, and Yihong Gong, "Locality-constrained linear coding for image classification.," in *CVPR*, 2010, pp. 3360–3367.

[20] Shenghua Gao, Ivor Wai-Hung Tsang, and Liang-Tien Chia, "Laplacian sparse coding, hypergraph laplacian sparse coding, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 92–104, 2013.

[21] Megha Pandey and Svetlana Lazebnik, "Scene recognition and weakly supervised object localization with deformable part-based models," in *ICCV*, 2011, pp. 1307–1314.

[22] Li-Jia Li, Hao Su, Eric P. Xing, and Fei-Fei Li, "Object bank: A high-level image representation for scene classification & semantic feature sparsification," in *NIPS*, 2010, pp. 1378–1386.

[23] Quannan Li, Jiajun Wu, and Zhuowen Tu, "Harvesting mid-level visual concepts from large-scale internet images," in *CVPR*, 2013, pp. 851–858.

[24] Jian Sun and Jean Ponce, "Learning discriminative part detectors for image classification and cosegmentation," in *ICCV*, 2013, pp. 3400–3407.

[25] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson, "Cnn features off-the-shelf: An astounding baseline for recognition," in *CVPR Workshops*, June 2014.

[26] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples," in *Workshop on Generative-Model Based Vision, IEEE Proc. CVPR*, 2004.

[27] Matthew D. Zeiler and Rob Fergus, "Visualizing and understanding convolutional networks," in *ECCV*, 2014, pp. 818–833.