Frog sound identification using extended $k$-nearest neighbor classifier

Nordiana Mukahar$^{1,2}$, Bakhtiar Affendi Rosdi$^1$, Dzati Athiar Ramli$^1$, Haryati Jaafar$^3$

$^1$Intelligent Biometric Group, School of Electrical & Electronic Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 Nibong Tebal, Pulau Pinang Malaysia
$^2$Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Malaysia
$^3$Faculty of Engineering Technology, Universiti Malaysia Perlis, UniCity Alam Sg. Chuchuh, 02100 Padang Besar, Perlis, Malaysia

E-mail: nordi741@tganu.uitm.edu.my, eebakhtiar@usm.my, dzati@usm.my, haryati.jaafar@unimap.edu.my

Abstract. Frog sound identification based on the vocalization becomes important for biological research and environmental monitoring. As a result, different types of feature extractions and classifiers have been employed to evaluate the accuracy of frog sound identification. This paper presents a frog sound identification with Extended $k$-Nearest Neighbor (EKNN) classifier. The EKNN classifier integrates the nearest neighbors and mutual sharing of neighborhood concepts, with the aims of improving the classification performance. It makes a prediction based on who are the nearest neighbors of the testing sample and who consider the testing sample as their nearest neighbors. In order to evaluate the classification performance in frog sound identification, the EKNN classifier is compared with competing classifier, $k$-Nearest Neighbor (KNN), Fuzzy $k$-Nearest Neighbor (FKNN) $k$-General Nearest Neighbor (KGNN) and Mutual $k$-Nearest Neighbor (MKNN) on the recorded sounds of 15 frog species obtained in Malaysia forest. The recorded sounds have been segmented using Short Time Energy and Short Time Average Zero Crossing Rate (STE+STAZCR), sinusoidal modeling (SM), manual and the combination of Energy (E) and Zero Crossing Rate (ZCR) (E+ZCR) while the features are extracted by Mel Frequency Cepstrum Coefficient (MFCC). The experimental results have shown that the EKNN classifier exhibits the best performance in terms of accuracy compared to the competing classifiers, KNN, FKNN, KGNN and MKNN for all cases.

1. Introduction
Frog as one of the unique amphibian with diverse benefits plays a central role in many ecosystems. The ecological role of frog is vital in the food chain, medical research and also serves as an indicator for an environmental health. Every different species of frog possess a unique sound as their sound can be received over a varying distance that allows and obstructive detection of their existence [1]. Frog sound recognition gain a lot of interest among researcher and several techniques including syllable segmentation [1], feature extraction and classification [2] have been developed in order to recognize the frog species based on their sound. A typical frog sound recognition system consists of five major stages including data collection, data preprocessing to improve the quality of data, signal segmentation, feature extraction and classification [3].
The KNN introduced by [4] is a classical non-parametric classifier that attracts interest among researchers in pattern classification and recognition. It measures the similarity between the testing sample and its neighbors using distance metric and the testing sample is assigned to the class that has a majority vote in \( k \) neighborhood. Conventional KNN treats each nearest neighbor with equal weight and the majority vote is applied in making decision. This can be a problem when the nearest neighbors vary widely in their distances and the closer neighbors more reliably predict the class of the testing sample [5]. Several weighting schemes have been proposed to resolve the ambiguity of the weighting distance between the testing sample and its nearest neighbors. Empirical works on frog sound recognition with MFCC as feature extraction and various types of a classifier such as Support Vector Machine (SVM), Sparse Representation Classifier (SRC) and Local Mean \( k \)-Nearest Neighbor with Fuzzy Distance Weighting (LMkNN-FDW) has been proposed in [3]. The fuzziness introduced in LMkNN-FDW [3] is based on fuzzy set theories that assign the fuzzy membership to the testing sample [6]. Another improvement of weighting scheme is proposed in [7] which the improved fuzzy is implemented with \( k \)-Nearest Centroid Neighbor (KNCN) classifier. In KNN classifier, most of the method proposed by the researcher are trying to resolve the drawbacks in KNN from one side perspectives only, testing sample. The judgment on deciding the nearest neighbors is made based on the which samples that their distance is close enough to the testing sample in feature space. In order to create fair and balanced treatment among neighbors, the mutual nearest neighborhood is proposed in [8] that defines one sample is considered as a mutual nearest neighbor only if both the testing sample and the training sample consider each other as their respective neighbor. The KGNN proposed in [9] is another variation of KNN that utilize both the neighborhood information of the testing sample and training sample in the decision rule. Unlike the MKNN that consider only the mutual neighborhood to measure the relation between the testing sample and its neighbors, the KGNN rule defines mutual neighborhood in broader perspective view. In order to decide whether the training sample is the general nearest neighbors of the testing sample, it takes into account the mutual neighborhood information of both kind of samples and the overlapping region of the mutual neighborhoods [9]. Both the MKNN and KGNN use a majority vote in decision rule.

In this paper, a new approach of utilizing the mutual neighborhood information of the testing sample and training sample to define the nearest neighbors is presented. It considers not only who the nearest neighbors of the testing sample are, but also who consider the testing sample as their nearest neighbors. The idea of a generalized class-wise statistic based on nearest neighbors was proposed in [5] to evaluate the relationship between two distribution data whether they are mixed well or widely spread. It represents the variation of data distribution from multiple classes, thus, the final classification process is according to the maximum gain of intra-class coherence.

2. The extended \( k \) nearest neighbor classifier

Differ from the conventional classifier that considers the nearest neighbors of the testing sample only, the EKNN defines the nearest neighbors by utilizing the neighborhood information of testing sample and training sample. Thus, the nearest neighbors are considered from two perspectives [5]: a) who is the neighbors of the testing sample, and b) who considers testing sample as their neighbors. The procedure of the proposed extended \( k \)-nearest neighbor classifier can be described as follows: let a set of training sample \( T = \{x_j \in \mathbb{R}^d\}_{j=1}^N \), with the label classes, \( c_1, c_2, ..., c_n \) where \( n \) is the number of classes and \( x_j \) is the training sample and \( N \) is the number of training sample in \( d \)-dimensional feature space.

- At training stage during offline, search the \( k \) nearest neighbors of each training sample from the training set, \( T \). The distance between training sample, \( x_j \) and its nearest neighbors, \( x_j^{NN} \), is measured by the Euclidean distance metric.
• Obtain the generalized class wise statistic for class $i$, $T_{i}^{NN}$.

$$T_{i}^{NN} = \frac{1}{n_{i}k} \sum_{x \in S_{i}} \sum_{r=1}^{k} I_{NN, r}(x, S = S_{1} \cup \ldots \cup S_{n})$$

(1)

$i = 1, 2, \ldots, n$

Where $S_{n}$ denotes the samples in class $n$, $x$ denotes one single sample in $S = S_{1} \cup S_{2} \cup \ldots \cup S_{n}$, $n_{i}$ denotes number of samples in $S_{i}$, and $k$ is the user defined parameter, number of nearest neighbors. The indicator function, $I_{NN, r}(x, S)$ indicates whether both the sample $x$ and $r$-th nearest neighbor belong to the same class. If both samples belong to the same class, then the outcome of indicator function equals to 1, otherwise, it equals to 0. This process can be done offline at training stage.

• During online with the introduction of the testing sample, $y$, find the $k$ nearest neighbors of a testing sample from the training set, $T$. Let $T_{NN}(y) = \{x_{j}^{NN} \in \mathbb{R}^{d} \}_{j=1}^{k}$ be the set of KNN for a testing sample, $y$. The similarity between the testing sample, $y$ and testing sample in feature space is measured using Euclidean distance.

• Investigate the total number of classes in the set of KNN, $T_{NN}$ for the testing sample, $y$. Let $m$ be the total number of classes to be investigated.

• Assume a testing sample, $y$ as the same class as the investigated classes and count the number of training sample who have an increased number of neighbors (from the same class) in its $k$ nearest neighbors (denote this number as $\Delta n_{NN}^{j}$). In this way, $\Delta n_{NN}^{j}$ represents the change of the $k$ nearest neighbors for class $i$ when a testing sample, $y$ is assumed to be class $j$.

• Calculate the generalized class wise statistic, $T_{NN, j}^{j}$ for class $i = 1, 2, \ldots, m$ when a testing sample, $y$ is assumed to be class $j$.

$$T_{NN, j}^{j} = \frac{1}{n_{i} \prime k} \sum_{x \in S_{i}, \prime j} \sum_{r=1}^{k} I_{NN, r}(x, S = S_{1} \cup \ldots \cup S_{n} \cup \{y\})$$

(2)

The parameter $n_{i} \prime$ is the size of $S_{i, \prime j}$ and it is defined as

$$S_{i, \prime j} = \begin{cases} S_{i} \cup \{y\}, & \text{when } j \leq i, \\ S_{i}, & \text{when } j \neq i, \end{cases}$$

(3)

• Finally, obtain the target function of EKNN, $f_{EKNN}$ which is the sum of generalized class wise statistic for each possible investigated class with equation 4 and assign the testing sample, $y$, to the class that has greatest target function among $m$ possible investigated classes according to equation 5.

$$f_{EKNN, j} = \sum_{i=1}^{m} T_{NN, j}^{i}$$

(4)

$$c = \arg \max_{j \in 1, 2, \ldots, m} \{f_{EKNN, j}\}$$

(5)

3. Experimental results

In order to evaluate the effectiveness of the proposed classifier, two experiments were conducted on frog call database from the Intelligent Biometric Group (IBG), Universiti Sains Malaysia (USM)[10]. The performance comparison is made in terms of classification accuracy with other
competing classifiers such as KNN [4], FKNN [6], KGNN [9] and MKNN [8]. Four types of
syllable segmentation technique involved for the performance comparison in both experiments
such as STE+STAZCR, SM, manual and E+ZCR. The total number of syllables used is 806 from
15 different species for the training and testing purpose. Both experiments were conducted using
10-fold cross validation strategy. Experiments have been carried out using MATLAB R2014a on
Intel(R) Core(TM) i7-4510U CPU processors running @ 2.00GHz and 2.60 GHz. The system
has 4 GB of memory and is running a 64-bit Windows 8 operating system.

3.1. Classification performance on different numbers of feature dimension for different syllables
segmentation technique
The first experiment intends to validate the classification performance between the EKNN
classifier and the competing classifiers on different sizes of feature dimension for different
syllables segmentation technique of frog sound. For the first experiment, the size of the
neighborhood, \( k \) is set equal to 5 while different sizes of feature dimension were compared
at 100, 256 and 1024. Figure 1 shows the classification accuracy in clean signal for various
classifiers in conjunction with four different types of syllable segmentation technique. It can
be observed from Figure 1 that the EKNN classifier consistently achieves the best performance
for each individual syllable segmentation compared to the KNN, FKNN, KGNN and MKNN.
The classification accuracy performance of the EKNN classifier is also superior to competing
classifiers for different values of feature dimension. The best classification accuracy of EKNN on
SM, manual, STE+STAZCR and E+ZCR syllable segmentation techniques are 95.28\%, 98.88\%,
98.51\% and 98.75\%, respectively.

3.2. Classification performance on different syllables segmentation technique with various sizes
of neighborhood, \( k \)
The second experiment is based on a comparison of classification performance between the
EKNN classifier and the competing classifier over different values of neighborhood size, \( k \). In
this experiment, the value of the neighborhood parameter \( k \) is set to an odd number ranging
from 1 to 15 while using the optimum values of feature dimension obtained from the previous
experiment for each syllable segmentation type as shown in Table 1. Figure 2 depicts the
comparison of classification accuracy of for different classifiers when \( k \) varies from 1 to 15. The
EKNN classifier performs better than other four classifiers among the interval neighborhood
size, \( k \) and the results of EKNN in the interval of \( k \) on E+ZCR, STE+STAZCR and manual are
slightly decreased before it stable with an increase of \( k \). It is noticeable that the classification
rates of KNN, FKNN, and KGNN are gradually decreased when the value of \( k \) is relatively large
and it is due to the facts that when the size of the neighborhood becomes larger, the nearest
neighbors are dominated by other class that led to misclassification problem. As opposed to
MKNN that exhibits the poor classification performance when the \( k \) value is small and slowly
rising with increasing of \( k \) because the larger the value of \( k \) is, the more mutual nearest neighbors.
The classification performance differences between EKNN and KNN, FKNN, KGNN, MKNN
are very significant especially for SM segmentation when the value of \( k \) becomes large.

4. Conclusion
In this paper, a new classification approach for a frog call recognition called Extended k-Nearest
Neighbor is presented. A new selection of nearest neighbors is based on maximum gain intraclass
coherence that utilizes the mutual neighborhood information for both testing and training
sample. In order to investigate the performance of classification accuracy, experiments were
conducted on frog sound database with 4 different types of syllable segmentation technique
such as SM, E+ZCR, STE+STAZCR and manual. The experimental results show that EKNN
Figure 1. Classification accuracy based on different sizes of feature dimension for different syllables segmentation.

Table 1. The optimum size of feature dimension for each syllable segmentation technique with respective classifier.

| Segmentation | KNN | KGNN | MKNN | FKNN | EKNN |
|--------------|-----|------|------|------|------|
| SM           | 100 | 100  | 100  | 100  | 100  |
| Manual       | 100 | 256  | 256  | 100  | 100  |
| STE          | 256 | 256  | 100  | 256  | 256  |
| E+ZCR        | 256 | 256  | 100  | 100  | 256  |

performs well and outperform the competing classifiers, KNN [4], FKNN [6], KGNN [9] and MKNN [8].
Figure 2. Classification accuracy for different syllable segmentation with parameter $k$ ranging from 1 to 15.

References

[1] Obrist M K et al. 2010 Abc Taza 8 68
[2] Huang C J et al. 2009 Expert Syst. Appl. 36 3737
[3] Jaafar H et al. (eds) 2014 Frog identification system based on local means $k$-nearest neighbors with fuzzy distance weighting (Springer)
[4] Cover T and Hart P 1967 IEEE T. Syst. Inform. Theory. 13 21
[5] Tang B and He H 2012 IEEE Comput. Intell. M 10 52
[6] Keller J M, Gray M R and Givens J A 1985 IEEE T. Syst. Man Cyb. 580
[7] Jaafar H et al. 2015 Comput. Intel. Neurose.
[8] Liu H and Zhang S J. Syst. Software 85 1067
[9] Pan Z, Wang Y and Ku W 2017 Knowl.-Based Syst. 121 142
[10] Jaafar H and Ramli D A 2015 J. Trop. Resour. Sustain. Sci 3 208