# MLANN; Maximum Likelihood Approximate Nearest Neighbor in Real-time Image Recognition

Ali Gholami
Department of Computer Engineering & Information Technology
Amirkabir University of Technology
https://aligholamee.github.io
aligholami7596@gmail.com

Abstract—In this report, the core algorithm of the MLANN is observed. We'll introduce the dataset which is going to be used as our train data. We'll also estimate the computational requirements for this project.

## I. Introduction

FOR the task of statistical image recognition, the unknown densities of each class should be estimated at the first step. Suppose we have R reference images as $r \epsilon \{1, 2, 3, ..., R\}$. Statistical image recognition can be reduced to computing the distances between the input image $X$ and the reference images $X_r$ which is formally illustrated below:

$$W_v : v = \arg\min_r \rho(X, X_r)$$

$W_v$ is the actual class of the input image.

## II. Initial Intuition

The core implementation of this algorithm is where the reference images are being selected. In the nearest neighbor method, the whole database is brute forced to find the nearest image to the input image. But in this method, there are conceptually different approaches that can speed up finding the best match for the input image. This method conducts the Best-bin-first kd-trees (An extension of classic kd-trees) to create a priority queue of the reference images. Next, the highest priority item $X_i$ is pulled from the queue and the set of reference images $X_i^{(M)}$ is determined by the following expression:

$$(\forall X_k \epsilon X_i^{(M)})(\forall X_j \not\epsilon X_i^{(M)})$$
$$|\rho_{i,k} - \rho(X, X_k)| \leq |\rho_{i,j} - \rho(X, X_j)|$$

Note that $\rho_{i,k}$ are the elements of the distance matrix $P = [\rho_{i,j}]$ which is computed in the preliminarily step.

## III. Assumptions

Here is the main assumptions of the algorithm we'll discuss below.

1) The reference images from different classes are independent.
2) The probability distributions of the feature vectors from the same class are identical.

Advisor: Professor Mohammad Rahmati

## IV. Core Algorithm

Let the reference images $X_{r_1}, X_{r_2}, ..., X_{r_k}$ have been checked before the k-th step, i.e. the distances $\rho(X, X_{r_1}), \rho(X, X_{r_2}), ...\rho(X, X_{r_k})$ are computed. The next reference image is selected using the following formula:

$$r_{k+1} = \underset{v \epsilon \{1,...,R\} - \{r_1,...,r_k\}}{argmax} (p_v. \prod_{i=1}^{k} f(\rho(X, X_{r_i})|W_v))$$

where $\prod_{i=1}^{k} f(\rho(X, X_{r_i})|W_v)$ is the conditional density of the distance $\rho(X, X_{r_i})$ if the hypothesis $W_v$ is true. If the prior probabilities of all classes are not identical, then the next instance will be selected from the majority classes with high probability.

## V. Datasets & Implementation Progress

The initial platform for the face recognition is implemented using OpenCV and Tensorflow. Faces are detected using the MTCNN[1] method which is an extension of the implementation of David Sandberg (FaceNet's Contributor). The features are extracted in real time using ResNet V1[2]. It is planned to generate a database of extracted features using 13000 images of the LFW dataset which consists of labeled faces in approximately 1600 categories. Note that the current implementation of the project, which we call it BioFace from this day forward, uses the naive nearest neighbor brute force search to find the best match of the input image among the reference images in database.

## References

[1] Zhang, Kaipeng, et al. "Joint face detection and alignment using multitask cascaded convolutional networks." IEEE Signal Processing Letters 23.10 (2016): 1499-1503.
[2] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.