# A supervised probabilistic model for visual object recognition

Oussama Aiadi
LINATI Laboratory
Univ Kasdi Merbah
Ouargla 30000, Algeria
oussama.aiadi@univ-ouargla.dz

Belal Khaldi
LINATI Laboratory
Univ Kasdi Merbah
Ouargla 30000, Algeria
khaldi.belal@univ-ouargla.dz

Mohammed Lamine Kherfi
LINATI Laboratory
Univ Kasdi Merbah
Ouargla 30000, Algeria
Lamine.kherfi@univ-ouargla.dz

Yacine Ghorfa
Univ Kasdi Merbah
Ouargla 30000, Algeria
yacine.gho@gmail.com

Rayhana Rezzag Bara
LMA Laboratory
Univ Kasdi Merbah
Ouargla 30000, Algeria
rihana.rezzag @gmail.com

***Abstract*—In this paper, we propose a new supervised probabilistic-based method for object recognition. Specifically, we conduct a process of supervised learning in which each class is represented by Gaussian Mixture Model (GMM). In order to group images having the same visual appearance, we cluster images related to each class using k-means. Probability density functions that correspond to the resulting clusters are fused in a GMM representing the class model, where Expectation-Maximization (EM) is used to estimate the parameters of the mixture. Given a test image, we calculate the probability that the image belongs to each class. The image is then assigned to the class having gained the highest score. The proposed method takes into account the intra-class variation and it is capable to distinguish different objects in spite of the small inter-class variation. Experimental results demonstrated the effectiveness of our method and an accuracy of 95.28% is reached.***

***Keywords*—Object recognition, Gaussian Mixture Model (GMM), supervised machine learning, web images, Expectation-Maximization (EM).**

## I. INTRODUCTION

Visual object recognition refers to the task of determining whether a particular object (i.e., object of interest) is present in the image. Regarding its high importance, object recognition is considered as one of the hottest topics of computer vision, as it covers a wide range of applications including image retrieval [1,2], industrial inspection [3,4] and robot navigation.

In the literature, several approaches have been proposed to deal with the object recognition issue. Multiple kernel learning (MKL) methods aim to pick out and combine kernels for a particular recognition task. MKL–based methods have been widely utilized for object recognition [5,6]. Other methods recognize objects by fusing features at different levels. As instance, in [7], shape and color features were fused to recognize different fruits. [8] proposed a weighted score-level feature fusion scheme for fruit recognition. Deep learning methods have been successfully applied for recognizing objects. [9] proposed an incremental object recognition method based on deep learning. Certain other methods [11,10,3] have focused their attention on recognizing a specialized type of objects. For instance, [11] proposed a method for underwater objects recognition via light detection and ranging systems. Nevertheless, none of those methods have taken into account the intra-class variations, as images belonging to the same class could have several visual appearances.

In this paper, we propose a new supervised probabilistic-based method for object recognition. In particular, our method consists in two stages namely training and testing. First, we describe images using different descriptors including HSV-based color histogram and Local Binary Patterns (LBP). As local information could be more discriminative, we partition images into blocks, and then we extract LBP features from each block. Afterwards, we use Principal Component Analysis to reduce the features dimensionality. Image that correspond to the same class are then clustered using the k-means algorithm in order to assemble visually similar images together. Images within each cluster are supposed to be normally distributed, thus, we associate each cluster with a Probability density function (pdf) of multivariate normal distribution. Functions related to the same class are combined in a Gaussian Mixture Model (GMM), where Expectation-Maximization (EM) algorithm is used for parameters estimation. Given a test image, we calculate the probability that the image belongs to each class. The image is assigned to the class having yielded the highest probability score.

The remainder of this paper is organized as follows. In Section 2, we detail the steps of our method. In Section 3, we present the experimental results we obtained. Finally, Section 4 draws some conclusions.

## II. PROPOSED METHOD

The steps of the proposed method can be summarized as follow:

- We denote images belonging to the class $C$ by $T_c$.

- *Features extraction:* in this step, features namely HSV-based color histogram and Local Binary Patterns (LBP) are extracted from ea$c$□ $T_c$.LBP features are extracted from each block of image.
- *Clustering:* we cluster $T_c$ into $K$ cluster, we get the set of clusters $Clust_c = \{Clust_{c_k}, k = 1, \dots, K\}$. Clustering is conducted using the k-means algorithm.
- *Modeling:* probability densities of clusters within each $Clust_c$ are combined in a GMM, such that parameters are estimated using Expectation- Maximization (EM) algorithm.
- *Test:* after calculating the probability that an image test belongs to each class, the image is assigned to the class having obtained the highest probability score.

Now, let us put more precisions about our method.

*II.1. Features extraction*

*II.1.1. Color features*

Color features are essential for retrieval and recognition tasks. In this work, we use the HSV-based color histogram as HSV color space is known to be closer to the human perception than the RGB space. For a compact yet efficient representation, the histogram is calculated over a quantized HSV space. The hue channel is quantized into 8 ranges, while the two remaining channels are quantized into 3 ranges. In addition, statistical moments namely mean and standard deviation are extracted from H and S channels.

*II.1.2. Local Binary Patterns (LBP)*

Besides color, object could also be distinguished based on texture features. Therefore, to boost the recognition yields, we combine the HSV color histogram with the LBP features. LBP is a powerful feature which has been extensively utilized in a wide range of applications [12] [13]. The basic LBP operator [12] is calculated by assigning each pixel with a binary code of eight bits. This code is computed by considering the center pixel of a 3x3 neighborhood as a threshold value. The appearance frequency of each code, called pattern, in the image is then counted to fill up histogram of 256 ( $2^8$) dimensions. In this work, we consider the uniform version of LBP which consists in 59 patterns. It is worth noting that each image is partitioned into no-overlapped blocks (25x25), then LBP features are extracted from each block. Finally, LBP features from different blocks are fused in a single vector.

*II.2. clustering*

We use the k-means to cluster each $T_c$ into $K$ clusters. We do so in order to assemble images having the same visual appearance within the same cluster. We empirically set the value of $k$ to 18.

*II.3. Concepts modeling using GMM and EM*

We represent each cluster in $Clust_c$ using the probability density function of M-dimensional normal distribution which is given by:

$$G\,(\boldsymbol{x}|\boldsymbol{\theta}_{\mathrm{C_k}}) = \frac{\mathbf{1}}{\sqrt{|\Sigma_k|\,(2\pi)^{M}}}\,\boldsymbol{e}^{-\frac{1}{2}(x-\mu_k)^{T}\,\Sigma^{-1}\,(x-\mu_k)}\,, \qquad (1)$$

where $x$ is a M-dimensional data point, $and$ $\theta_{\mathrm{C_k}}$ denotes the parameters of the Gaussian distribution that corresponds to the $k^{th}$ cluster, these parameters are:

$\mu_k$: The mean vector of the points belonging to the cluster.
$\sum_k$: The covariance matrix.

Probability density functions (pdf) that correspond to the set of clusters $Clust_c = \{Clust_{c_k}, k = 1, \dots, K\}$ are combined in a Gaussian Mixture Model (GMM) which is given by:

$$\boldsymbol{P_c}\,(\boldsymbol{x}|\boldsymbol{\theta_c}) = \sum_{k=1}^{k_c} \boldsymbol{w_{c_k}}\, \boldsymbol{G}\,(\boldsymbol{x}/\boldsymbol{\theta}_{c_{\mathrm{k}}})\,,\ \boldsymbol{\theta_{cp}} = \{\boldsymbol{\theta_{C_k}}, \boldsymbol{k} = \mathbf{1}, \dots, \boldsymbol{K}\}, \qquad (2)$$

where $w_{c_k}$ denotes the weight of the $k^{th}$ distribution of a class $C$ and $\theta_c$ the parameters of Gaussian component densities of the mixture $P_c$.

The likelihood of the data is given by:

$$\mathrm{L}=\prod_{i=1}^{N_c} P_c\,(x_i|\theta_c), \qquad (3)$$

where $N_c$ denotes the number of images in the class $C$.

Expectation-Maximization (EM) algorithm is then used to maximize L and estimate the parameters of $\theta_c$.

1. Expectation step:

$$y_{ij} = \frac{\frac{w_j}{\sqrt{|\Sigma|\,(2\pi)^{M}}}\, e^{-\frac{1}{2}(x_i-\mu_j)^{T}\,\Sigma^{-1}\,(x_i-\mu_j)}}{\sum_{i=1}^{N_c} y_{ij}} \qquad (4)$$

2. Maximization step:

$$w_j = \frac{\sum_{i=1}^{N_c} y_{ij}}{N_c}, \qquad (5)$$

$$\mu_j = \frac{\sum_{i=1}^{N_c} x_i\, y_{ij}}{\sum_{i=1}^{N_c} y_{ij}}, \qquad (6)$$

$$\Sigma_j = \frac{\sum_{i=1}^{N_c} y_{ij}\ (x_i - \mu_j)(x_i - \mu_j)^T}{\sum_{i=1}^{N_c} y_{ij}}, \qquad (7)$$

These two steps are repeated for $i= 1,\ldots, N_c$ and $j= 1,\ldots, K$ until convergence.

*II.4. Image recognition*

During test, each test image $I$ is associated with a score, denoted as $S_{c,}$ representing their probability of appertaining to the class $C$. $I$ is assigned to the class having the highest score.

## III. EXPERIMENTAL SETUP

*III.1. Dataset*

In this work, experiments were conducted on the fruit dataset that is introduced by [8]. The dataset is made up of 6610 images from 20 classes including orange, tomato, Banana…etc. 50 images from each class are intended for testing, and the remaining for training. For more details about the dataset, reader can refer to [8].

*III.2. Performance metrics*

We consider the recognition accuracy as a performance measure. The recognition accuracy for a single class is defined by

$$Accuracy_v = \frac{Number\ of\ correctly\ recognized\ samples}{Number\ of\ samples\ assigned\ to\ the\ variety} \times 100$$

The average accuracy is given by

$$Accuracy_o = \frac{\sum_1^N Accuracy_v}{N},$$

where $N$ is the number of classes.

## IV. EXPERIMENTAL RESULTS

In order to examine the efficiency of the proposed method, we conduct experiments on the fruit dataset. First, we test the recognition capability of each of the used features. Then, we test the strength of the different possible combinations. Table 1 shows the average accuracy yielded by each combination. From Table 1, we note that single color features have yielded low accuracies (26.66% and 33.86%). In addition, combining the two features haven't improved the accuracy. This may be attributed to the strong visual resemblance between certain images that belong to different classes. We can note also that LBP alone has reached an accuracy of 90.26% which confirms its discrimination capabilities. In addition, we observe that combining each of the color features with the LBP has yielded a high accuracies (93.67% and 92.20% ). By combining all the features, an accuracy of 95.28% has been reached.

TABLE 1: average accuracy yielded by the different combinations

| | Accuracy |
|---|---|
| LBP | 90.26 |
| HSV color histogram | 26.66% |
| Statistical moments | 33.86% |
| LBP + HSV color histogram | 93.67% |
| LBP + Statistical moments | 92.20% |
| HSV color histogram + Statistical moments | 28.66% |
| HSV color histogram + Statistical moments + LBP | **95.28%** |

To confirm the strength of the proposed method, we compare it with certain state of art methods. In particular, we make comparison with two different methods for object recognition. In the first one [8], a weighted combination scheme was proposed for the task of fruit recognition. In the second one [10], a genetic algorithm-based fusion weight selection method is proposed to recognize multisensory activity of Elderly people. Table 2 shows the average accuracy yielded by our method and the competing methods.

TABLE 2: average accuracy yielded by the different methods

| Method | Accuracy |
|---|---|
| The method in [10] | 89% |
| The method in [8] | 90.7% |
| The proposed method | **95.28%** |

From Table 2, we can note that the proposed method significantly outperformed the competing method in terms of accuracy.

## V. CONCLUSION

In this paper, we have presented a new method for visual object recognition. We have used a GMM to represent each class. As images in the same class could have different visual appearances, the proposed method takes into consideration this intra-class variation. In addition, it is capable to distinguish the highly resembled images from different classes. Further improvements could improve the performance of our method. For instance, combination of the GMM with other classifiers using different combination schemes. Experimental results have proved the effectiveness and the efficiency of our method.

## VII. REFERENCES


[1] B. Khaldi and M. L. Kherfi, “Modified integrative color intensity cooccurrence matrix for texture image representation,” Journal of Electronic Imaging, vol, 25, no. 5, 2016.

[2] M.L. Kherfi, D. Ziou, and A.Bernardi, “Atlas WISE: A Web-based image retrieval engine”, In Proceedings of the International Conference on Image and Signal Processing(ICISP), pp. 69-77, 2003.

[3] O. Aiadi and M.L. Kherfi, “A new method for automatic date fruit classification”, International Journal of Computational Vision and Robotics vol, 7, no. 6, 2017.

[4] O. Aiadi and M.L. Kherfi, “An automated system for date fruit recognition through images”, Revue des BioRessources, vol. 6, no. 1, 2016.

[5] H. Hoashi, T. Joutou, and K. Yanai, “Image recognition of 85 food categories by feature fusion,” in IEEE Int. Symp. on Multimedia, pp. 296–301 (2010)

[6] P. Gehler and S. Nowozin, “On feature combination for multiclass object classification,” in IEEE 12th Int. Conf. on Computer Vision, pp. 221–228 (2009)

[7] A. Harjoko and A. Abdullah, “A fruit classification method based on shapes and color features,” in 3rd Asian Physics Symp., pp. 445–448(2009)

[8] H. kuang, L. L. H. Chan, C. Liu and H. Yan, “Fruit classification based on weighted score-level feature fusion”, Journal of Electronic Imaging, vol, 25, No. 1,2016.

[9] L. Yan, Y. Wang, T. Song, Z. Yin, “An incremental intelligent object recognition system based on deep learning”in Chinese Automation Congress, 2017.

[10] S. Chernbumroong, S. Cang, and H. Yu, “Genetic algorithm-based classifiers fusion for multisensor activity recognition of elderly people,” IEEE J. Biomed. Health Inf. 19(1), 282–289 (2014).

[11] S. Matteoli, G. Corsini, M. Diani, G. Cecchi, G. Toci, “Automated underwater object recognition by means of fluorescence LIDAR”, IEEE Transactions on Geoscience and Remote Sensing, vol, 53, No. 1,2015..

[12] M. Topi, O. Timo, P. Matti, S. Maricor; “Robust texture classification by subsets of local binary patterns”, in Proc. of International Conference on Pattern Recognition, Vol. 3, pp. 935–938, (2000)

[13] T. Ahonen, A. Hadid, M. Pietikäinen; “Face Description with Local Binary Patterns: Application to Face Recognition”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 12, pp. 2037–2041, (2006).