Indoor Scene Recognition based on Weighted Voting Schemes

Alejandra C. Hernandez[1], Clara Gomez[1], Erik Derner[1,2,3] and Ramon Barber[1]

***Abstract*— Scene understanding represents one of the most primary problems in computer vision. It implies the full knowledge of all the elements of the environment and the comprehension of the relationships between them. One of the major tasks in this process is the scene recognition, on which we focus in this work. Scene recognition is a relevant and helpful task in many robotic fields such as navigation, localization, manipulation, among others. The knowledge of the place (e.g. "office", "classroom" or "kitchen") can improve the performance of robots in indoor environments. This task can be difficult because of the variability, ambiguity, illumination changes, occlusions and scale variability present in this type of spaces. Commonly, this problem has been approached through the development of models based on local and global characteristics, incorporating context information and, more recently, using deep learning techniques. In this paper, we propose a multi-classifier model for scene recognition considering as priors the outcomes of independent base classifiers. We implement a weighted voting scheme based on genetic algorithms for the combination of different classifiers in order to improve the recognition performance. The results have proved the validity of our approach and how the proper combination of independent classifier models makes it possible to find a better and more efficient solution for the scene recognition problem.**

***Index Terms*— Scene recognition, multi-classification, scene understanding, mobile robots, genetic algorithms.**

I. INTRODUCTION

Scene recognition is a challenging problem in computer vision. It consists in classification of a given scene to one of the specified categories. To achieve this, it is necessary to analyze geometric and semantic information as well as the relationships between the scene elements to develop robust scene recognition models. Depending on the type of place and the elements present in it, the classification process can be easier or more challenging. Several conditions such as variability, ambiguity, illumination, viewpoint and scale changes, among others, affect the process of identifying a scene. Due to the wide range of applications in which it is used such as autonomous robots, intelligent robotics and human-robot interaction, this field is of major importance in mobile robotics. The categorization of a scene is a complex task that implies the understanding of the content of the scene and the relations and interactions of its content. For example, consider these two scene categories: a meeting room and a dining room. Both categories contain similar objects such as chairs and tables. In these cases, the identification of objects is not sufficient to distinguish the scenes, as the relationships and the arrangement of these objects are essential to obtain a better prediction of the scene. For this reason, scene recognition is a more challenging problem than other image recognition tasks.

Different techniques have been applied to deal with scene recognition. Some of them consider different strategies such as handcrafted image features [1], learning features [2], incorporating contextual information [3] and others are based on more complex learning methods such as convolutional neural networks [4]. Despite the promising results obtained so far, the problem of scene recognition is still far from being definitively solved.

The main motivation of this work is to profit from the advantages that different techniques can offer and merge them into a multi-classifier to obtain a more robust model for scene recognition. In this paper, we present a multi-classifier model for the task of indoor scene recognition based on weighted voting schemes. A multi-classifier, also known as ensemble classifier or combined classifier, makes it possible to build a more robust model that combines the results of independent base classifiers. We propose two weighted voting schemes, one based on the performance rate of the independent classifiers and another one based on genetic algorithms to compute the weights for each base classifier to obtain a final estimation of the place where the robot is considering what it can perceive. Genetic algorithms calculate the optimal weights that are assigned to individual classifiers in the multi-classifier in order to obtain satisfactory results. Figure 1 shows the scene recognition outputs for two scene categories, a living room and a laboratory of a university.

The main contributions of this work are:

- The combination of different techniques for independent base classifiers in order to improve the scene recognition result. This way, the proposed model can adequately merge different features as well as object information and any other contextual information.
- A more general and scalable multi-classifier model for scene recognition that can be adjusted to any type of source information.
- A compensation of errors made by each base classifier fusing them through a weighted combination of their classification contributes to make the best estimate of the scene perceived by the robot.

The remainder of the paper is structured as follows: Section II presents the related research in the field of scene recognition. Section III describes the proposed model for

[1]Robotics Lab, Department of Systems Engineering and Automation, Carlos III University of Madrid, Spain.

[2]Czech Institute of Informatics, Robotics, and Cybernetics, Czech Technical University in Prague, Czech Republic.

[3]Department of Control Engineering, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic.

Corresponding author: Alejandra Hernandez, alejhern@ing.uc3m.es.

978-1-7281-3605-9/19/$31.00

Fig. 1: Example of the results of the proposed multi-classifier model. Two different scenes can be observed, a living room and a laboratory. The model fuses the outcomes of the base classifiers through a weighted voting scheme to give the final prediction about the place that the robot perceives.

scene recognition based on weighted voting schemes. The base classification models used in this work are presented in Section IV. In Section V the two weighted voting schemes are explained. Section VI describes the results and finally, Section VII draws conclusions and outlines future work.

II. RELATED WORK

Scene recognition is one of the fundamental problems in robotics which has been investigated for decades. This task is one of the most important for scene understanding. The information about the scene, its elements and the relationships between them can improve the comprehension of the environment where the robot operates and also directly influences the performance of the robot in human-robot and robot-environment interaction tasks. Great efforts have been made to develop accurate models for scene recognition. Some approaches are based on handcrafted image features. In [1] the authors proposed to use generic and specific characteristics to address the problem of scene classification. Different features are considered: geometric features, size and shape, pixel depth, geocentric pose and the appearance of superpixels. In [5] SIFT features for all the images (grayscale and color) using dense extraction methods are obtained. Then, a bag of features approach is used in order to generate a training model. In [6] sparse codes of SIFT features as local appearance descriptors are used to implement a variation of the spatial pyramid matching method for scene classification.

Other works use the information of the objects to improve the scene recognition rate. In [3] scene and object information are combined through convolutional neural networks. A single CNN is used as feature extractor. They trained the network with different datasets with images of scenes and objects with different scale ranges. In [7] common objects are associated with a class of scene considering contextual information. To do that, a search strategy to find meaningful objects in the scene is implemented. Then, the found objects are combined through a Naive Bayes approximation to predict the scene class.

More recent works are focused on convolutional neural networks (CNN) that are trained on large image datasets to solve the scene recognition problem. In [8] RGB and depth features are combined in a multi-modal fusion framework for scene recognition. Authors in [4] proposed a CNN based on 3D information to obtain a representation of the environment. The training set consists of 3D scenes templates that include the furniture arrangement. In this way, it is possible to include the context information in the model.

Single-classifier approaches have lead to decent and accurate results in some situations, however, it is still far from having a classifier that succeeds in every environment and situation. For this reason, our approach is to use a combination of classifiers where limitations are overcome by taking advantage of the strengths of individual classifier models. Multi-classifier models, also known in the literature as combined classifiers or ensemble classifiers, involve the combination of heterogeneous or homogeneous classifiers to generate a final decision [9]. The combination of classifiers has been extensively used to improve the classification results in face recognition [10], [11], emotion identification [12], texture recognition [13] and object detection [14]. This is not the case of scene recognition where few approaches including multi-classifiers have been presented.

In [15], authors propose SVM ensembles to address the classification of rare scenes. Rare scenes are those whose negative class is much larger than the positive class in standard datasets. A hierarchical SVM strategy is presented in this work. Each SVM is trained independently and the results are aggregated with another SVM. An approach for scene classification by creating an ensemble of parallel deep rule-based classifiers is proposed in [16]. Each classifier is trained separately using the features obtained from a pre-trained DCNN. The decision layer is designed as a 'winner-takes-all' that considers the result for each classifier and the confidence scores. In [17] a single CNN classifier is used. Classification is performed independently based on the features for different layers of the CNN. Outputs of different layers are combined to perform scene categorization. A soft combination of the ensemble of layers of the classifier is proposed that uses static and dynamic weights computed with genetic algorithms. Similarly, in [18], a CNN with a Local Convolutional Supervision (LCS) layer is used. LCS keeps the importance of fine-grained and detailed information in the image improving the final result for scene recognition. Important local information is encoded in a Fisher Convolutional Vector (FCV), which compensates the high-level FC-features for scene classification.

Unlike the approaches mentioned above, in this work independent base classifier models that extract different features from the images are combined to give a prediction of the scene where the robot is located. The classification outputs are fused using dynamic weighted voting schemes through two strategies, considering the performance rate of each independent base classifier and using genetic algorithms.

III. METHOD OVERVIEW

The combination of several classifiers can lead to a meaningful improvement in the recognition rate and the overall system performance. The development of a multi-classifier model requires several base classifiers first. Each one of these classifiers can be designed using several techniques, input features and different sources of information. Each base classifier generates an output with the prediction about the class to which a scene belongs. Then, these outputs must be appropriately combined in order to obtain a final prediction.

In this paper, we propose a multi-classifier model for scene recognition. The structure of the model is based on parallel topology. Each base classifier receives RGB and depth images. They perform the scene classification independently according to their feature extraction methods and classification algorithms. The outputs of the classifiers are the probabilities of the classes to which the scene belongs. The outputs of the classifiers are combined using a weighted voting scheme. Figure 2 shows a general overview of the proposed model.

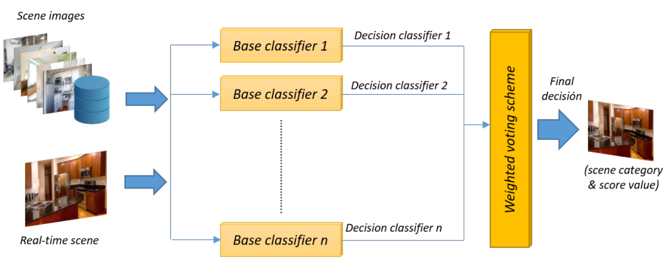

Fig. 2: General overview of the proposed multi-classifier model for scene recognition. The final decision of the combination of the outputs of the base classifiers is based on weighted voting schemes.

The proposed model is capable of working with an arbitrary number of base classifiers. As an example, for this work, two base classifiers are used. The first base classifier model consists of a scene recognition system based on machine learning that generates as output the score values for each scene category. On the other hand, a probabilistic scene recognition model that incorporates the information about the objects is used as another base classifier. This second model, developed in [19] gives as final result the class and the probability of the scene perceived by the robot. Each base classifier generates a ranked list of the classes with their likelihood values.

To ensure an adequate combination of the results of the base classifiers we have decided to implement two weighted voting schemes. In the first voting scheme, the weights are assigned to each classification output considering the individual performance of each scene recognition model. The accuracy of each model is used to determine the final result about the scene perceived by the robot. The second scheme makes use of genetic algorithms to determine the optimal weights that have to be applied to the output of each base classifier model. Finally both voting schemes are compared to determine the best solution for the scene recognition problem.

IV. Base Scene Recognition Models

To build a multi-classifier model, several base classifiers are necessary. In this work, we use two models as base classifiers. First, a scene recognition system based on machine learning is employed. The model makes use of images of different indoor environments to generate a training model. Color and grayscale images are used as input data. The model is divided in two stages, training stage and classification stage. For feature extraction, the technique of Bag of Features (BoF) combined with SURF descriptors are employed. A classification method based on Multilayer Perceptron is used. When deployed in a real environment the robot identifies the current scene according to the trained model and the result is a list of the trained classes with their respective score values. Figure 3 shows a general diagram of the first base classifier.

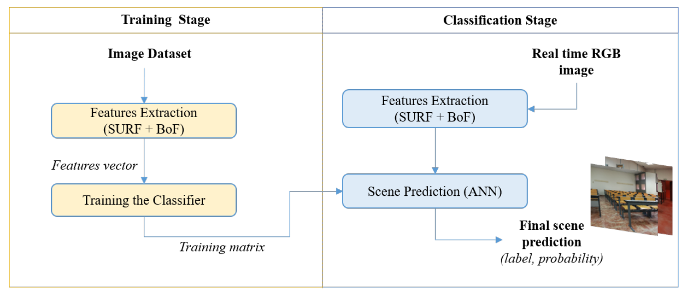

Fig. 3: A general representation of the scene recognition model based on BoF and SURF descriptors. The model uses Multilayer perceptron as classification method.

As a second base classifier, a probabilistic scene recognition model that considers multiple objects in the scene as prior information is used. The model developed in [19] employs support vector machine (SVM) as a classification method and considers the relationship between the objects in the scene to determine the final result about the class of the scene that is perceived. Figure 4 shows a representation of the second base classifier model. For the training process, two available datasets have been used, the SUN397 scene recognition benchmark [20] and the Kyushu University Indoor Semantic Place Dataset [21].

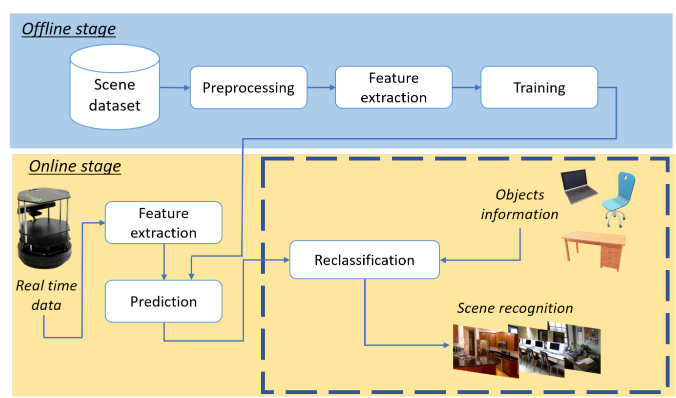

Fig. 4: Probabilistic Scene Recognition Model consisting of two stages: offline and online. The reclassification process considers scene and object relations to refine the final probability.

To include the influence of several objects in the prediction process, we have implemented the Bayes Extended Theorem that allows to combine multiple conditions with independent ancestors. We incorporate the information about not only one object at a time but several objects at the same time. The probability of being in a scene ξ_k given several objects at the same time $\cap_s O_s$ is shown in Equation 1.

$$P(\xi_k | \cap_s O_s) = \frac{P(\cap_s O_s|\xi_k) * P(\xi_k)}{P(\cap_s O_s)} \quad (1)$$

where the probability of detecting several object in a scene, $P(\cap_s O_s|\xi_k)$, is defined as shown in Equation 2.

$$P(\cap_s O_s|\xi_k) = \prod_s P(O_s|\xi_k) \quad (2)$$

These probabilities were obtained through co-occurrence matrices that indicate how often the objects appear in a specific scene. $P(\xi_k)$ corresponds to the initial probability of being in a scene. Finally, $P(\cap_s O_s)$ is the probability to find several specific objects in any scene category.

V. WEIGHTED VOTING SCHEMES

The selection of appropriate weights is crucial to have a robust multi-classifier model. In this section, two weighted voting scheme strategies are studied.

A. Weighted Voting Scheme by Accuracy

In this type of voting scheme a weight is assigned to each base classifier model. The weights are based on the assumption that classifier models with high recognition performance are more reliable than classifiers with lower performance rate. For this work, the weight of each base classifier is represented by the accuracy of each model. The accuracy is a measure to evaluate the performance of a classification model that represents the proportion of the correct predictions that is determined using the Equation 3.

$$Accuracy = \frac{N^{\mathrm{o}}\, of\, correct\, predicted\, images}{Total\, number\, of\, images} \quad (3)$$

The process consists of multiplying each scene class probability obtained from each base classifier model by its respective weight. Next, the weights of the classifiers are added for each class of scene. Then, the final prediction is the scene class with the highest sum of weights. Equation 4 shows the formulation of this voting scheme:

$$\hat{y} = \underset{i}{argmax} \sum_{j=1}^{m} \omega_j P_{ij} \quad (4)$$

where $\hat{y}$ represents the final result after the combination process, ω_j is the weight of each base classifier j associated to the accuracy of the model and finally, P_{ij} is the prediction result of the base classifier j for the scene class i.

B. Weighted Voting Scheme through Genetic Algorithms

In this section, a second voting scheme for the combination of base classifiers is proposed. This voting scheme can be considered more general than the voting scheme based on the performance of the system. First, a set of weights as free parameters are considered. Then, the idea is to find a combination of values that generates the best result for the whole system. There are many optimization methods that can be applied. In this work, we have decided to use genetic algorithms because of the robustness and good performance demonstrated in many other complex problems [22].

Genetic Algorithms (GAs) [23] are a subset of the computer science branch called Evolutionary Computation. GAs are search-based algorithms inspired by the processes of natural and genetic selection. The process begins with a group or population of possible solutions to a given optimization problem. Each candidate solution is called a chromosome and initially each of them is created randomly. It is necessary to define a fitness function to evaluate the quality of the solution. Each chromosome is assigned a fitness value, and only the most fit individuals, that is, the chromosomes that constitute a better solution to the optimization problem, are allowed to 'reproduce'. During reproduction, new individuals are created from the random changes (mutation) and fusion of two chromosomes (crossover). This way, better individuals or solutions are obtained over the generations (evolving), until a stop criterion is reached.

The phases of a GA are: initialization, fitness function definition, selection, crossover and mutation. The details of each phase with respect to our method are described below.

1) Initialization: Each chromosome is represented by an array of weights that are real numbers between 0 and 1. The number of elements of the array is equal to the number of base classifier models (m). All the values of the chromosomes are generated randomly. For this work, the population size (N) is set to 50.

2) Fitness Function: The fitness function or objective function $F_{obj}(i)$ is defined as the accuracy of each base classifier model (A_n) combined with the weights (ω_n) corresponding to each chromosome i. Through this function, shown in Equation 5 a fitness score for each chromosome is calculated to determine if an individual is able to compete with others.

$$F_{obj}(i) = |\sum_{n=1}^{m} (A_n * \omega_n) - 1| \quad (5)$$

3) Selection: In this step the idea is to select the fittest individuals. The fitness of each chromosome is calculated according to Equation 6.

$$Fitness(i) = \frac{1}{1 + F_{obj}(i)} \quad (6)$$

The selection probability is calculated considering the fitness score of each chromosome (Equation 7). Lower values of $Fitness(i)$ lead to lower selection probabilities.

$$P_{selection}(i) = \frac{Fitness(i)}{\sum_{i=1}^{k} Fitness(i)} \tag{7}$$

4) Crossover: This is the most important step in a genetic algorithm. During the crossover, for each pair of parents that must be mated, a random crossover point is chosen within the genes. This way, new population is created by fusing the information in the chromosomes.

5) Mutation: This step involves replacing a gene in a random position with a new value. This process guarantees the diversity of the population and avoids early convergence. Once the mutation process is done, the iteration of the GA has finished, yielding a new generation. With this, the objective function is evaluated. If the result of the objective function decreases, it means that a better solution was found compared with the previous chromosome. Finally, the best chromosome represents the final result of the weights that will be applied for the fusion of the base classifiers.

VI. EXPERIMENTAL RESULTS

A. Experimental Setup

The proposed multi-classifier model for indoor scene recognition has been designed to run on a real mobile robot. For the experiments, we have used a Turtlebot 2 robotic platform equipped with an RGB-D sensor for the recognition of objects and scenes. To demonstrate the usefulness of our approach in real scenes, two environments have been selected: a building of a university and a typical house. In all the experiments the robot was teleoperated and the different modules of the proposed multi-classifier model operated autonomously.

B. Evaluation of the Weighted Voting Scheme by Accuracy

The first experiment consisted of moving the robot by teleoperation in the two selected environments, capturing different scenes, while the base classifiers were executed at the same time. Each of the base classifiers yields the probability of the scene of belonging to each of the classes for which they were designed. The multi-classifier model receives this information and fuses it using as weights the accuracy of each model obtained during its respective training phase. This accuracy values are 78.5% for the first base classifier model and 88.4% for the second base classifier model. Table I shows the final results of the base classifier models and the result of the proposed multi-classifier model using a weighted voting scheme based on accuracy. The first two columns show the recognition rate of each base classifier model (69.98% and 76.48% respectively). The last column shows the recognition rate of the proposed multi-classifier model (80.17%) and how the scene recognition task improves after merging the base classifier models considering the accuracy as weights.

C. Evaluation of the Weighted Voting Scheme through Genetic Algorithms

The second experiment was carried out under the same conditions as the first one. The robot moves through the two selected environments, a university and a typical house, classifying the scene that it perceives. The voting scheme strategy used in this experiment is based on genetic algorithms. Figure 5 shows the prediction results in a scene type 'kitchen'. Table II shows the configuration of the main parameters used in the generation of the weights through genetic algorithms.

TABLE I: Evaluations of the multi-classifier model based on accuracy. Environments: university building and typical house

Evaluations	Recog. rate 1	Recog. rate 2	Final Recog. rate
Laboratory	72.45%	79.86%	87.58%
Garage	63.97%	74.62%	78.76%
Classroom	72.01%	72.32%	74.79%
Univ. Avg	69.48%	75.60%	80.38%
Kitchen	78.77%	89.20%	91.47%
Bedroom	72.93%	78.99%	79.96%
Bathroom	63.22%	76.05%	78.90%
Living room	67.01%	65.20%	69.54%
House Avg	70.48%	77.36%	79.97%
Total Avg	69.98%	76.48%	80.17%

TABLE II: Main parameters used in the implementation of genetic algorithms for scene recognition.

Parameters	Values
Population size	50
Nº iterations	100
Crossover rate	0.8
Mutation rate	0.1

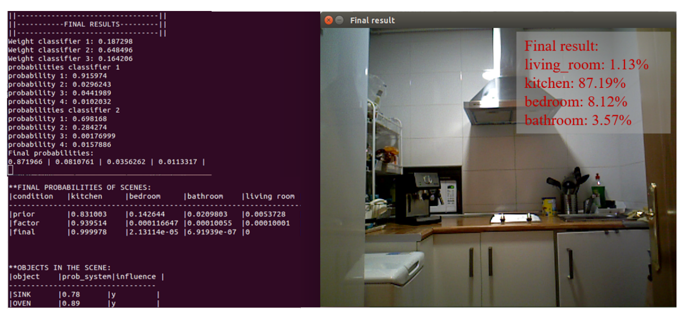

Fig. 5: Execution of the multi-classifier model based on weighted voting scheme through GA in a common house. The information of each base classifier model are considered to obtain the final result.

After the weights are generated, they are multiplied by the individual results of each base classifier model. In table III the results of the application of our multi-classifier model based on GAs are presented.

The results show that the implementation of the two weighted voting schemes proposed in this work increases the recognition rate in comparison with each of the base classifier models. The voting scheme based on accuracy achieved good results with a 80.17% of average recognition rate. However,

TABLE III: Evaluations of the multi-classifier model based on genetic algorithms.

Evaluations	Recog. rate 1	Recog. rate 2	Final Recog rate
Laboratory	71.86%	77.86%	87.72%
Garage	69.42%	81,06%	86.36%
Classroom	70.91%	75.44%	79.52%
Univ. Avg	70.73%	78.12%	84.54%
Kitchen	82.95%	91.78%	95.94%
Bedroom	73.77%	78.15%	85.60%
Bathroom	57.38%	72.85%	75.43%
Living room	66.56%	67.88%	73.07%
House Avg	70.16%	77.66%	82.51%
Total Avg	70.45%	77.89%	83.52%

weighted voting scheme based on genetic algorithms produces the best result for the scene recognition task with an average recognition rate of 83.52%. The experiments show that the proper combination of independent classifiers can compensate for errors during the classification process and generate much more robust recognition models.

VII. CONCLUSIONS

We proposed a multi-classifier model for scene recognition based on weighted voting schemes. In this work, two strategies of weighted voting schemes were considered. In the first strategy, the weights assigned to each base classifier were equal to the respective accuracy of each model. On the other hand, in the second scheme, a genetic algorithm was implemented for weight optimization. In this process the accuracy was used to create the objective function.

The experiments show that the recognition rate using the proposed multi-classifier is better than using individual classifiers. In this work, we have implemented two variants of voting schemes that achieve both good results. However, the weighted voting scheme based on genetic algorithms works better than the simple voting scheme by accuracy. The adequate combination of independent classifiers through genetic algorithms allows to obtain a more robust and precise model for scene recognition, taking advantage of the benefits of each classifier and compensating the errors of each of them. As future work, we plan to study other voting strategies, increase the number of base classifiers and incorporate a new base classifier based on deep learning techniques in order to evaluate the robustness and performance in a multi-classifier model.

VIII. ACKNOWLEDGEMENTS

This research has received funding from HEROITEA: Heterogeneous Intelligent Multi-Robot Team for Assistance of Elderly People (RTI2018-095599-B-C21), funded by Spanish Ministerio de Economía y Competitividad. This work was also supported by the European Regional Development Fund under the project Robotics for Industry 4.0 (reg. no. CZ.02.1.01/0.0/0.0/15_003/0000470) and by the Grant Agency of the Czech Technical University in Prague, grant no. SGS19/174/OHK3/3T/13.

REFERENCES

[1] S. Gupta, P. Arbelaez, and J. Malik, “Perceptual organization and recognition of indoor scenes from rgb-d images,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 564–571.

[2] H. Yin, X. Jiao, Y. Chai, and B. Fang, “Scene classification based on single-layer sae and svm,” *Expert Systems with Applications*, vol. 42, no. 7, pp. 3368–3380, 2015.

[3] L. Herranz, S. Jiang, and X. Li, “Scene recognition with cnns: objects, scales and dataset bias,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 571–579.

[4] Y. Zhang, M. Bai, P. Kohli, S. Izadi, and J. Xiao, “Deepcontext: Context-encoding neural pathways for 3d holistic scene understanding,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1192–1201.

[5] L. Zhou, Z. Zhou, and D. Hu, “Scene classification using a multi-resolution bag-of-features model,” *Pattern Recognition*, vol. 46, no. 1, pp. 424–433, 2013.

[6] J. Yang, K. Yu, Y. Gong, T. S. Huang *et al.*, “Linear spatial pyramid matching using sparse coding for image classification.” in *CVPR*, vol. 1, no. 2, 2009, p. 6.

[7] P. Espinace, T. Kollar, N. Roy, and A. Soto, “Indoor scene recognition through object detection using adaptive objects search,” in *European Conf. on Comp. Vision, Workshop on Vision for Cognitive Tasks*, 2010.

[8] H. Zhu, J.-B. Weibel, and S. Lu, “Discriminative multi-modal feature fusion for rgbd indoor scene recognition,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2016, pp. 2969–2976.

[9] M. Woźniak, M. Graña, and E. Corchado, “A survey of multiple classifier systems as hybrid systems,” *Information Fusion*, vol. 16, pp. 3–17, 2014.

[10] H. M. Raafat, A. S. Tolba, and A. M. Aly, “A novel training weighted ensemble (twe) with application to face recognition,” *Applied Soft Computing*, vol. 11, no. 4, pp. 3608–3617, 2011.

[11] L. Nanni, A. Lumini, and S. Brahnam, “Ensemble of texture descriptors for face recognition obtained by varying feature transforms and preprocessing approaches,” *Applied Soft Computing*, vol. 61, pp. 8–16, 2017.

[12] M. Bejani, D. Gharavian, and N. M. Charkari, “Audiovisual emotion recognition using anova feature selection method and multi-classifier neural networks,” *Neural Computing and Applications*, vol. 24, no. 2, pp. 399–412, 2014.

[13] F. Yang, G.-S. Xia, G. Liu, L. Zhang, and X. Huang, “Dynamic texture recognition by aggregating spatial and temporal features via ensemble svms,” *Neurocomputing*, vol. 173, pp. 1310–1321, 2016.

[14] T. Malisiewicz, A. Gupta, A. A. Efros *et al.*, “Ensemble of exemplar-svms for object detection and beyond.” in *Iccv*, vol. 1, no. 2. Citeseer, 2011, p. 6.

[15] R. Yan, Y. Liu, R. Jin, and A. Hauptmann, “On predicting rare classes with svm ensembles in scene classification,” in *Int. Conf. on Acoustics, Speech, and Signal Processing.*, vol. 3. IEEE, 2003, pp. III–21.

[16] X. Gu, P. P. Angelov, C. Zhang, and P. M. Atkinson, “A massively parallel deep rule-based ensemble classifier for remote sensing scenes,” *IEEE Geosc. and Rem. Sensing Lett.*, vol. 15, no. 3, pp. 345–349, 2018.

[17] S. Bai and H. Tang, “Softly combining an ensemble of classifiers learned from a single convolutional neural network for scene categorization,” *Applied Soft Computing*, vol. 67, pp. 183–196, 2018.

[18] S. Guo, W. Huang, L. Wang, and Y. Qiao, “Locally supervised deep hybrid model for scene recognition,” *IEEE transactions on image processing*, vol. 26, no. 2, pp. 808–820, 2017.

[19] A. C. Hernández, C. Gómez, R. Barber, and O. Martínez Mozos, “Object-based probabilistic place recognition for indoor human environments,” in *2018 International Conf. on Control, Artificial Intelligence, Robotics Optimization (ICCAIRO)*, 2018, pp. 177–182.

[20] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, “Sun database: Large-scale scene recognition from abbey to zoo,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 3485–3492.

[21] O. M. Mozos, H. Mizutani, H. Jung, R. Kurazume, and T. Hasegawa, “Categorization of indoor places by combining local binary pattern histograms of range and reflectance data from laser range finders,” *Advanced Robotics*, vol. 27, no. 18, pp. 1455–1464, 2013.

[22] O. Kramer, *Genetic algorithm essentials*. Springer, 2017, vol. 679.

[23] J. H. Holland *et al.*, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.