ABSTRACT

"Indoor scene recognition", also called "scene classification and scene identification", refers to the process of labeling the elements of the given input scene image based on the contents. Basically there are two approaches to design an Indoor Scene Recognition System (ISRS): i) Feature-based and ii) Object-based. Due to its wide applications, scene recognition has gained great research attention over the past decennial. Even though different methods have been proposed in the literature, there is no consensus on the type of classification in a more prefect manner. Also performance of the scene recognition systems is found to be less when compared with the process involved in it. Hence in this research work two indoor scene recognition systems are designed based on features of the given scene and objects of the scene that can provide higher performance.

This research work initially proposes a new novel ISRS based on features of the given scene image. These features are extracted from the low level primitives, namely homogeneity, edge and texture present in the scene image. To extract these features, the proposed system utilizes the well-known orthogonal polynomials model. A new block decomposition model is designed in the transformed domain with orthogonal polynomials. The polynomials effects and mean square amplitude responses are computed and the features are extracted with inherent feature selection process on each block of the input scene image under consideration. The novelty of the proposed feature extraction is its reduced dimensionality. The extracted features are then fed to the SVM classifier for the purpose of classifying the scene. The proposed feature-based ISRS could produce a higher accuracy of 83.88%.

This research work also proposes a new object-based ISRS instead of considering the low level features of the scene image. In this direction, the proposed object-based ISRS introduces three novel cost effective algorithms. The proposed system introduces the object detection as an intermediate representation. At first, the given scene image is subjected to object detection and recognition with well-known CNN based YOLO v3 algorithm. These identified objects are then divided into sets of mandatory and desirable objects with a simple look-up manner based on a set of possible objects and their classification, designed specifically for the proposed system, with indoor environment clubbed with human knowledge on standard datasets. The proposed system introduces a new *scene-object identification* phase for the purpose of designing a possible scene that these mandatory and desirable objects may point to, with a new SOI algorithm. Also a new binary scene representation (BSR) is proposed for the purpose of finalizing the scene classification by computing a new score called *scene-score* with the BSR algorithm. The performance of the proposed object-based ISRS is experimented on standard datasets and the proposed system could produce a higher accuracy of 98.27%. The performance of the proposed object-based ISRS is also compared with existing similar systems and the results are encouraging.

With an objective of improving the performance of the above designed object-based ISRS, this research work introduces a new object classification technique based on simple CNN architecture. This architecture takes the detected and recognized objects present in the scene under analysis with YOLO v4 algorithm. The aim of the proposed CNN architecture is to classify the recognized objects into a set of mandatory and desirable objects. This CNN architecture consists of four convolution and max-pooling layers, besides a flatten layer, dense layers and output layer for the purpose of proposed object classification. The novelty of this proposed CNN architecture

is to identify non-significant objects, besides natural classification of mandatory and desirable objects. This identification of non-significant objects greatly reduces the time complexity involved in the scene recognition.

The proposed object-based ISRS also introduces another CNN architecture for the purpose of object characterization. This new CNN architecture makes use of higher number of filters in convolution and max-pooling layers, in addition to increased dense layers so as to facilitate the proposed scene characterization. As a result of scene characterization, the system produces the possible scenes. The novelty of this part of research work is the introduction of newly modified CNN architecture that could replace the previously introduced *scene-object* classification technique with look-up table mechanism. The proposed refined object-based ISRS in terms of CNN architecture for object characterization then utilizes already defined BSR coding technique for the purpose of finalizing the scene recognition. The efficiency of the proposed CNN based object classification and modified CNN architecture for object characterization are experimented with standard benchmark datasets and the performances are measured with standard metrics.