Situation, Scene and Scenario Classifications and Understanding

Dr. Jaiprakash Narain Dwivedi

Kolar Road, Bhopal, India-462042

ABSTRACT

The study of Situation, Scene and Scenario Classification is the necessary because of the understanding of basic techniques towards the recognition of level of risk of danger especially during unpredictable infinite possibility of situations. To infer high-level semantic scene categories, Scene Classification is useful and important step. It becomes necessary especially in the case of low-level visual features. A basic challenge for natural situation, scene and scenario classification is due to the fast growing researches in the field of autonomous driving.

Keywords: Situation, Scene and Scenario

I INTRODUCTION

According to Cambridge dictionary [2], the set of things that are happening and the conditions that exist at a particular time and place, the position of something, especially a town, building, etc. The paper [3] defines the entirety of circumstances, which are to be considered for the selection of an appropriate behavior pattern at a particular point of time. It entails all relevant conditions, options and determinants for behavior.

According to Cambridge dictionary [46], scene is the place where a unit of action or some event occurs, any views or picture, display of anger, strong feeling, bad manners, a division of a play or an act of a play, an incident or situation in real life, etc. As per definition of paper [16] Scene is a place in which one can have movement. In order to understand, we further study the scene in detail.

According to Cambridge dictionary [44], a description of possible actions or events in the future, a written plan of the characters and events in a play or film. The functional description of driver assistance systems and also in the context of simulation and testing, the term scenario" is found generally.

In the section2, situation study, in the section3, hierarchical classification and hierarchical situation classification, in the section4, scene study and its category has been explained. Section5 and section6, describes hierarchical scene classification and scenario study respectively and finally section7 is for conclusion.

II SITUATION STUDY

A situation is derived from the scene by an information selection and augmentation process based on transient (e.g. mission-specific) as well as permanent goals and values. Hence, a situation is always subjective by representing an element's point of view."

2.1 Situation of road-vehicles

The situation of road and vehicles on the road are unpredictable. The situation of road- vehicles is being explained with the figure 1 [1].

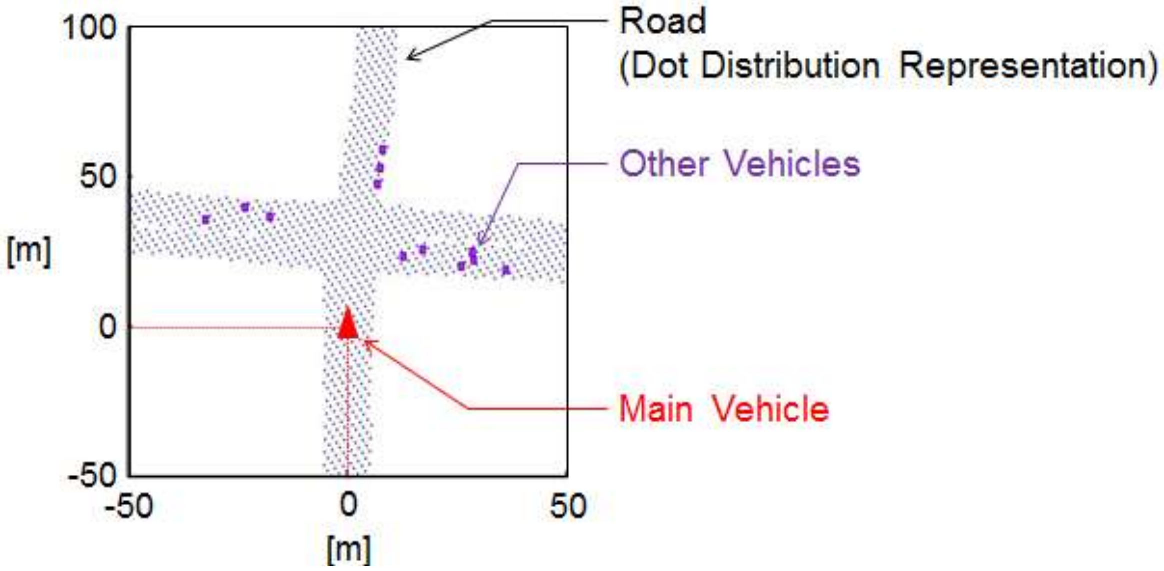

Figure 1: Representation of Situation of road-vehicles. The different possible position of the vehicles are represented using magenta color, main vehicle in red color and the shape of the road using dot distribution representation.

Situation consists of the pair of road and vehicles as shown in figure 1. In this figure, the shape of road is being represented using dot distribution representation. The positions of other vehicles are represented using magenta color in square shape. The triangular shape in red color is being used to show the position of main/subject/ego-vehicle at the origin. The x and y axes are measured in meter.

III HIERARCHICAL CLASSIFICATION

Hierarchical structures exist on all levels of natural scenes. A multiple instance learning by maximizing diverse density can be used to classify images of natural scenes [47].

As we have seen the scene categorization in the SUN dataset [27] and the term outdoor manmade and this further categorized as many different scenes. The first level of hierarchical study, classify the number of different kinds of the shapes of the road. The second level of the hierarchy is the number of objects on these different shapes of the road. The benefits of hierarchical study are listed below.

To make object, image and scene recognition easy.

To reduce the calculation cost and time.

To make fast retrieval of information.

To preserve the natural constrains and maintain simplicity.

Clarify of routes to follow the final groupings.

3.1 Hierarchical Situation Classification

Hierarchical Situation Classification has been shown in figure 2 [1]. This figure shows the example of hierarchy of road and the different positions of vehicles on the road. In the first level of hierarchy five different kinds of shapes of the road are considered. Three different shapes of the road of each category of the shape of the road have been shown in the level first of the hierarchy. In the second level of hierarchical study the number of different vehicles shown on the corresponding shapes of the road.

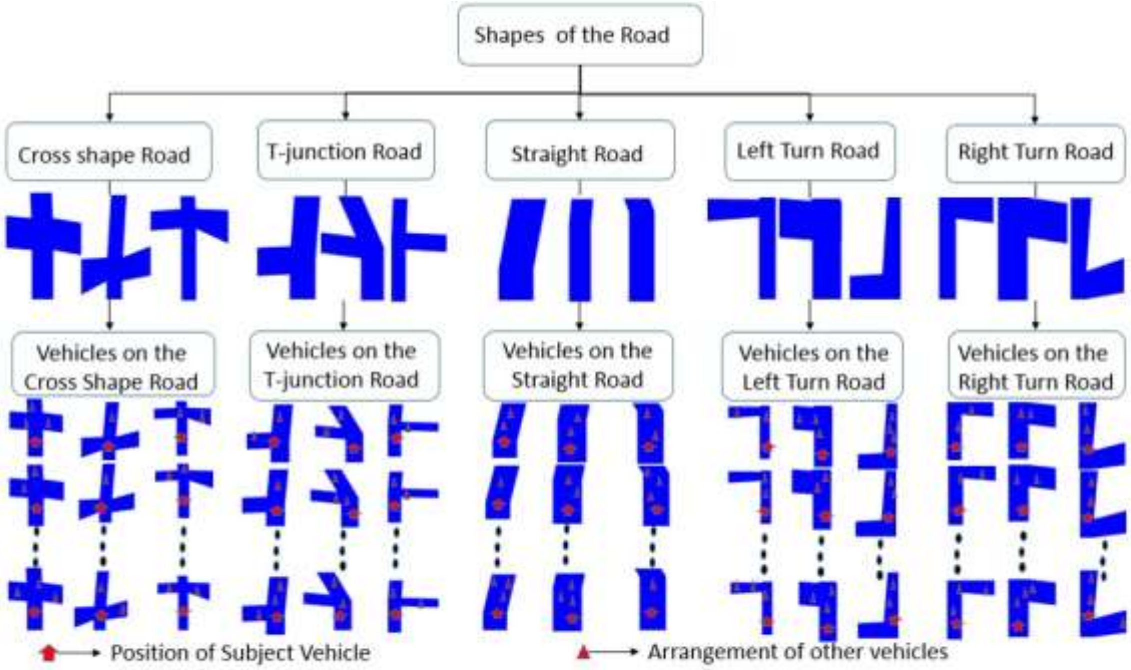
Shapes of the Road
Cross shape Road
T-junction Road
Straight Road
Left Turn Road
Right Turn Road
Vehicles on the Cross Shape Road
Vehicles on the T-junction Road
Vehicles on the Straight Road
Vehicles on the Left Turn Road
Vehicles on the Right Turn Road
Position of Subject Vehicle
Arrangement of other vehicles

Figure 2: This figure shows the example of hierarchy road and vehicles on it. In the first level of hierarchy five different kinds of shapes of the road are considered. In the second level of hierarchy the number of different vehicles shown on the corresponding shapes of the road.

IV RELATED WORK

For the situation assessment, sensors and database or digital map has been used as an input. Situation assessment thus helps to risk assessment and decision making system in a collision avoidance system [4]. The consideration of situation is with obstacle detection during movement of robot [5]. A learning classification algorithm to produce internal concepts is termed as situations, which can be used for constructing a graph of situation areas i.e. situation map of free space. When the robot moves the sensors collect these data and classify these data and thus situations are recognized and situation maps are formed [6]. If people are easily distracted, how do we maintain their situational awareness and how is their attention grabbed when it is needed. Situations in which inattention of driver (e.g., due to multitasking) might have led to an accident [7].The information of the current driving situation to establish a situation dependent data distribution. By using, a situation dependent data distribution the bus load in a vehicle can be reduced significant [8]. The evaluation of situation analysis in a common road scenario on a German highway. The recognition rates of about 89% of the driving situations in this scenario and by using the current driving situation for data distribution the bus-load could be reduced by about 17% [9]. In the model of situation, information of in-car sensors has been used to build up a representation of the environment around the ego vehicle. A situation analysis is established to detect the current driving situation according to the given description of driving situation on top of the situation model. Situation analysis is used to establish relationships (not necessarily hierarchical) and associations among entities in the situation model, it should anticipate with a priori knowledge in order to rapidly gather, assess, interpret and predict what these relationships might be; [10]. As autonomous vehicles advance toward handling realistic road traffic, they face street scenarios where the dynamics of other traffic participants must be considered explicitly. These situations include everyday driving maneuvers like merging into traffic flow, passing with on- coming traffic, changing lanes, or avoiding other vehicles [11]. The active safety systems has to operate properly even in hardly foreseeable traffic situations [12]. Situation assessment in tactical behavior planning for lane changes, whether lane changes are beneficial and/or possible [13]. A variety of different driving situations has been discussed. The different weather situations has been considered for navigation systems [14].

V STUDY

The study of scene in three divisions as small-scale scene, medium scale scene and large scale scene is furnished as below. These divisions are based according to datasets used for the study of scenes. These datasets have many common properties such as follows. One, they are iconic and typical scene images that background objects; consist of surfaces to decide scene categories. Two, these datasets have very limited number of scene categories and within the same class; images have very few variations in visual patterns. Three, they focus more on usually-seen categories, such as coast, forest and living room, etc. and so on.

5.1 Small-Scale Scenes Study

These small scale scenes are further divide into two categories for our easy understanding the scene. First one is outdoor scene category and second one is indoor scene category. Mixed category such as outdoor and indoor scene category is also studied in this study.

5.1.1 Outdoor Scene Category

This scene category uses the dataset of 8-scene dataset [15], [30], in which the image descriptor GIST was proposed. It consists of 8 outdoor scene categories in which 4 of them are from Natural landscapes and 4 of them are from man-made scenes. Totally, there are 2688 images in the dataset. It is popular benchmark in the scene understanding field and a mandatory challenge for all scene understanding researches. Later, researchers from the same group labeled the dataset for the purposes of semantic segmentation in scene parsing researches.

5.1.2 Outdoor and Indoor Scene Category

To create more challenging datasets, researchers intended to increase the categories and diversities of scenes. After the popularity of 8-scene dataset [15], 15-scene dataset has been introduced including 2 extra outdoor categories and 5 indoor categories to meet the requirement of scene diversities tasks. The 15-scene dataset [17], [19] and [30] has three more challenging factors than the above stated 8-scene datasets as follow.

(1) The 15-scene dataset consists of only gray-scale images.

(2) The 15-scene dataset has twice number of scene categories and images.

(3) The 15-scene dataset considers indoor scenes with outdoor scenes together.

Being such differences and challenges, the 15-scene dataset also became a mandatory benchmark in nowadays researches.

5.2 Medium-Scale Scenes Study

This medium scale scenes study describes the details about the event centric scene category and ground truth geometric levels category with the UIUC sports datasets and CMU 300 datasets respectively.

5.2.1 Event centric scene category

The dataset used for the study of event centric scene is UIUC sports [18], [30]. It consists of 8 sports event categories as follow. (i) rowing (250 images), (ii) badminton (200 images), (iii) polo (182 images), (iv) bocce (137 images), (v) snowboarding (190 images), (vi) croquet (236 images), (vii) sailing (190 images), and (viii) rock climbing (194 images). Each image has provided the information of the distance of the foreground objects. According to the human subject judgment, images are divided into easy and medium group. This dataset is prepared based on event happened during sports, therefore names event-centric scenes. The most scene images have distinctive foreground and background contexts. Hence, a promising research direction in context based recognition has been arisen using scene/object topic models.

5.2.2 Ground Truth geometric levels scene category

The largest benchmarking dataset in the geometric layout research community is dataset in [20], consists of 300 images of outdoor scenes with 23 different scene categories including building, alley, college, cliff etc. The dataset provides ground truth geometric labels for each image, namely support, sky, planar left, planar center, planar right, non-planar solid and porous. The first 50 images are used for training the surface segmentation and the remaining 250 images are used for evaluation. It provides occlusion boundaries for 100 images in the dataset other than geometric labels. The occlusion boundaries indicate occluded objects and the depth orders of occluding.

5.3 Large scale scene study

Due to big progresses in computer vision researches, small and medium scales datasets were no longer sufficient to meet the urgent need of large scale datasets and also for the evaluation purpose of robust scene understanding system performance.

5.3.1 Pascal Visual Object Classes (VOC) Challenge dataset

There are 20 object classes in the PASCAL dataset [24], [30] with thousands of images in each class. The PASCAL Visual Object Classes (VOC) Challenge dataset provides a common set of tools to access the data sets with annotations and standardized image datasets for recognition and object classification. This PASCAL VOC Challenge dataset was an annually arranged event from 2005 to 2012, in which researchers submitted their results on object classification, and got their results evaluated and compared online. The source of images was from the collection of Flicker photos. The online competition consists of classification, detection, segmentation, person layout and classification of action as follow.

Classification: To determine the presence or absence of an example of that class in the test image for each of the 20 classes.

Detection: To determine the label of each object and the bounding box from the 20 target classes in the test image.

Segmentation: To generate the pixel-wise segmentation of an object in the image.

Person layout: To determine the label and the bounding box of each part of a person.

Classification of Action: To determine the action of a person in a still image.

5.3.2 80 Million Tiny Image dataset

The dataset consists of 7,527,697 images, named as 80 million tiny image proposed by [21], [30]. The motivation for the preparation of this dataset was an interesting experimental observation i.e. human can achieve a recognition rate higher than 80% and also classify a scene with 32 × 32 pixels [21]. Each image is labeled with one of the 53,464 English nouns from the WordNet [22] and of low resolution (with image size dimension 32 × 32). This tiny image dataset is mainly used in fast image search which demands very little memory. Also, the main aim of this dataset is to facilitate the development of fast image search and scene matching techniques with very little memory. The sources of all images were the Google search and other engines using English nouns from the WordNet. These images contain object images and scene images are with high diversity. Its combination includes Caltech 101 categories, Caltech 256 categories, Vogel and Schiele [23] 702 natural scenes, Olivia and Torralba's [15] 2688 images. In object recognition task, The Caltech 101 categories and Caltech 256 categories images containing objects are used at a large scale. This proposed dataset contains image categories of sufficient diversity. In spite of large number of images in this dataset, all images are categorized. The confidence map, labels provided by users, nouns from the WordNet and the visual dictionary view of the whole dataset available in the project website [21]. The algorithmic accuracy in the classification of different categories is reflected by this confidence map. The amount of data provided by users is shown by labels. The averaged version of a class of images can reflect the global information of this class and the visual dictionary view supports the visualization of tiles by averaging the color of images that have the same English noun. The website provides the facility to add annotations by selecting a word and indicating whether correct and wrong images are returned to the users. To solve inaccurate annotations, the website allows users to correct labels online. However, these tiny images with low resolutions usually do not allow image classification algorithms to work properly.

5.3.3 LabelMe dataset

The dataset LabelMe [26], [30] is dynamic, free to use, and open to public contribution and prepared by the MIT CSAIL with an objective to provide a dataset of digital images with object and surfaces annotations. This LabelMe dataset provides a website to annotate images online for the users and asks people to use polygons to segment and annotate object and surfaces in an image. The following aspects separates this LabelMe are from other existing datasets.

(1) LabelMe contains images of objects with multiple angles, sizes and orientations.

(2) LabelMe designs images for object recognition in arbitrary scenes and it avoids the scene to cropped,

normalized or resized.

(3) Each image in LabelMe may contain more than one object, and users are allowed to label these objects.

(4) Its numbers of images and object classes can be easily increased.

Also to fix the following shortcoming this project was conducted. Most available data in computer vision research are tailored to the problem of a specific research group and it is often that new researchers need to collect additional data to solve their own problems.

5.3.4 Hierarchical ImageNet dataset

The hierarchical ImageNet dataset [25], [30] is an image dataset organized according to the Word Net hierarchy. There are more than 100,000 synsets in WordNet, and a great majority of them are nouns (80,000+). Each meaningful concept in the WordNet, possibly described by multiple words or word phrases, is called a "synonym set" or "synset". Compared to the other image classification dataset, the ImageNet is the largest and most challenging dataset for object classification and recognition. The ImageNet aims to provide on average 1000 images to illustrate each synset. This ImageNet dataset is mostly foreground object oriented like eivent centric scene category, UIUC sport dataset. Images of each concept are quality-controlled and human-annotated. On the other hand, it focuses on general image classification challenges, which include scene classification as only a small branch of the problem.

5.3.5 The SUN dataset

The SUN dataset [27] contains 899 categories and 130,519 images and proposed by Xiao et al., finds applications in many research fields, such as scene recognition, computer vision, human perception, cognition and neuroscience, machine learning, data mining, computer graphics and robotics research. In the SUN database images are all about scenes where human can navigate or interact with, which makes this dataset separate from the object detection datasets such as the PASCAL [24] and the Caltech 256 category datasets [28]. The motivation of this SUN dataset was to build a rich and diverse dataset that includes our daily experienced scenes in the real world as much as possible. As compared with the 80 million tiny image dataset, images in the SUN dataset are of much higher resolution. Images in the SUN dataset have a resolution of at least 200 × 200 pixels. Degenerate or unusual images (black and white, distorted colors, very blurry or noisy, incorrectly rotated, aerial views, noticeable borders) were removed in the image collection process. The SUN dataset is currently the largest scene dataset in terms of the image number and the number of scene. The scene category of the SUN dataset is huge. We can easily think of some scene categories such as the coast, the field, the meeting room etc. Is Grand Canyon a scene? Should it be a category? How can we include as many scene terms as possible? The category terms are selected from the 70,000 terms of the WordNet [22] used in the tiny images dataset [21]. These terms describe scenes, places and environments. There are several criteria in selecting scene category terms. First, places terms that are too broad to evoke a specific visual identity (such as territory, workplace and outdoors) and places names (such as Grand Canyon or NewYork) are not included. Second, specific types of objects which are scene related are included, such as buildings (skyscraper, house and hangar) which makes the scene categories more diverse. Third, it contains specific domains such as the pine forest, rainforest and orchard which all belong to the wooded area. The SUN dataset really contains comprehensive scene categories as shown in figure 2.1.

5.3.6. Places205 dataset

Place205 [29] is the latest and the largest scene classification dataset, and mainly prepared for the purposes of Convolution Neural Network (CNN) training. It contains 2,448,873 images from 205 scene categories. In Comparison with ImageNet and SUN, shows extreme data abundance, which is very crucial for discriminative model learning for CNN with deep structures that has millions of parameters [30]. In paper [29], author trained the huge CNN network with 2,448,873 image in 6 days and present superior results on traditional datasets with the trained deep features. Comparison of the numbers of images in Places 205 with ImageNet and SUN. Note that ImageNet only has 128 of the 205 categories, while SUN contains all of them. To compare them, we select a subset of Places. It contains the 88 common categories with ImageNet such that there are at least 1000 images in ImageNet. We call the corresponding subsets SUN 88 and ImageNet 88.

VI RELATED WORK

A weakly supervised feature learning method has been proposed to learn a discriminative and shareable filter bank to transform local image patches into features. After combining these features with the ConvNets features pre-trained on ImageNet, state-of-the-art scene classification result has been achieved [31].

A scene classification is being achieved using object distribution. A probabilistic Latent Semantic Analysis (PLSA), a generative model from the statistical text literature has applied to a bag of visual words representation for each image of scenes [32].

A scene classification task is performed using Spatial Pyramid Matching (SPM) and Hierarchical Dirichlet Processes (HDP) [33]. SVM ensembles has been used for scene classification for the rare class problems and experimental evidences shows that hierarchical SVM method performs better results in comparison to other techniques like majority voting, sum rule, neural network gater [34].

Sparse features derived from performing independent components analysis (ICA) on the power spectrum of images are more effective and more efficient (with fewer number of features) in classifying images into natural and manmade classes, compared to PCA type of features [35].

A framework has been presented to handle the problem of semantic scene classification, where a natural scene may contain multiple objects such that the scene can be described by multiple class labels. Cross-training is more efficient in using training data and more effective in classifying multi-label data. C-Criterion using threshold selected by MAP principle is effective for multi-label classification. Alpha-Evaluation, our novel generic evaluation metric, provides a way to evaluate multi label classification results in a wide variety of settings [36]. A visual grammar has been described to bridge the gap between low-level features and high-level semantic interpretation of images. Naive Bayesian classifiers has been used to learn models for region segmentation and classification from automatic fusion of features, fuzzy modeling of region spatial relationships to describe high-level user concepts, and Bayesian classifiers to learn image classes based on automatic selection of distinguishing (e.g., frequently occurring, rarely occurring) relations between regions. The visual grammar overcomes the limitations of traditional region or scene level image analysis algorithms which assume that the regions or scenes consist of uniform pixel feature distributions [37].

A patch-based latent variable model tailored for semantic scene classification tasks, where a latent layer of variables are used to model high level latent contextual visual concepts that are both predictable from the low-level feature inputs and discriminative for the semantic output labels. The proposed model can capture both local information, through the patches, and global information, through the summarization of the latent representation vectors in the whole image and the spatial regularization across patches, for target semantic label prediction [38]. A joint video scene segmentation and classification based on Hidden Markov Model (HMM) has been proposed. The "two-pass" approaches are based on the likelihood results from fixed-Length observation segments whereas "one-pass" approach uses a concatenating super-HMM network. The implementations of the "two-pass" approaches are simpler but they are less accurate in locating the scene transitions, because transitions are identified at the segment level. The "one-pass" approach is more accurate in identifying scene transitions because it makes decision at the clip level, but for the same reason, it can lead to a noisier segmentation. The proposed approaches not only segment video and classify them simultaneously, but also eliminate some of those false classification resulted from a simpler segment-based maximum likelihood approach because the proposed approaches take the neighboring information into consideration [39]. Bag-of-visual-word is an effective image representation in the classification task, but various representation choices w.r.t its dimension, weighting, and word selection has not been thoroughly examined. Techniques used in text categorization, including term weighting, stop word removal, feature selection, to generate various visual-word representations, and studied their impact to classification performance on the TRECVID and PASCAL collections [40]. An approach to holistic scene understanding that reasons jointly about regions, location, class and spatial extent of objects, presence of a class in the image, as well as the scene type. Learning and inference are efficient at the segment level, and introduce auxiliary variables that allow us to decompose the inherent high-order potentials into pairwise potentials between a few variables with small number of states. Inference is done via a convergent message-passing algorithm, which, unlike graph-cuts inference, has no sub -modularity restrictions and does not require potential specific moves [41]. Paced learning of Exemplar SVMs has been used to solve the problem of simultaneously learning a part model and detecting its occurrences in the training data. The distinctiveness of parts has measured by the new concept of entropy-rank, capturing the idea that parts are at the same time predictive of certain object categories but shareable between different categories. The learned parts have been shown to perform very well on the task of scene classification, where they improved a very solid bag of words or Fisher Vector baseline that in itself establishes the new state-of-the-art on the MIT Scene 67 benchmark [42].

6.1 Hierarchical Scene Classification (HSC)

Figure 3 [1] has been shown to make variety of scene in generalized hierarchy. In this figure, scenes have been classified on the basis of natural and manmade i.e. artificial scene. Further classification has been done in terms of indoor and outdoor scenes.

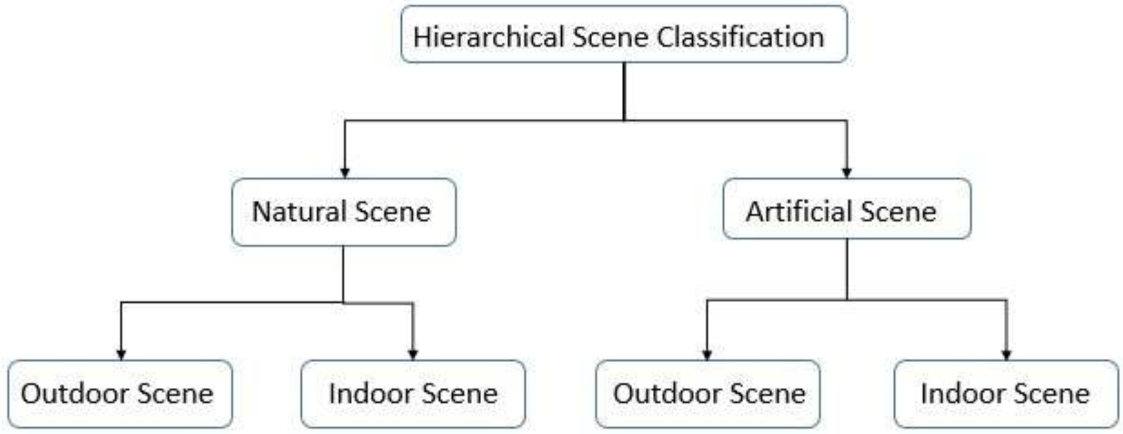

Figure 3: This figure shows the hierarchical classification of scenes.

6.2 Related work

A scene categorization based on bag of visual words representation has been presented. The classic approach is augmented by computing it on sub regions defined by three different hierarchically subdivision schemes and properly weighting the Textons distributions with respect to the involved sub regions. The weighed bags of visual words representation is coupled with a Support Vector Machine to perform classification. A similarity distance based on Bhattacharyya coefficient is used together with a k-nearest neighbor to retrieve scenes. Despite its simplicity, the proposed method has shown promising results with respect to state of the art methods [47]. A feed-forward convolutional network demonstrates and trained end-to-end in a supervised manner, and fed with raw pixels from large patches over multiple scales, can produce state of the art performance on standard scene parsing datasets. The model does not rely on engineered features, and uses purely supervised training from fully-labeled images to learn appropriate low-level and mid-level features. The pixel-wise accuracy is a somewhat inaccurate measure of the visual and practical quality of the result. Spotting rare objects is often more important than accurately labeling every boundary pixel of the sky (which are often in greater number). The average per-class accuracy is a step in the right direction, but not the ultimate solution: one would prefer a system that correctly spots every object or region, while giving an approximate boundary to a system that produces accurate boundaries for large regions (sky, road, grass), but fail to spot small objects. A reflection is needed on the best ways to measure the accuracy of scene labeling systems. Scene parsing datasets also need better labels. One could imagine using scene parsing datasets with hierarchical labels, so that a window within a building would be labeled as “building” and “window”. Using this kind of labeling in conjunction with graph structures on sets of labels that contain is-part-of relationships would likely produce more consistent interpretations of the whole scene. A pixel labeling system for training the convolutional net in isolation from the post processing module that ensures the consistency of the labeling and its proper registration with the image regions [48]. Gradients can be back propagated through the post-processor to the convolutional nets. This is reminiscent of the Graph Transformer Network model, a kind of non-linear CRF in which an un-normalized graphical model based post-processing module was trained jointly with a convolutional network for handwriting recognition. A more importantly advantage of joint training would allow the use of weakly-labeled images in which only a list of objects present in the image would be given, perhaps tagged with approximate positions. This would be similar in spirit to sentence-level discriminative training methods used in speech recognition and handwriting recognition [49].

Based on manifold learning to learn a hierarchical image manifold for Web image classification, assuming the images in an image set are usually related to the same or similar object but with various scenes. To achieve good classification performance and effectively reduce the computational complexity, a coarse-to-fine processing strategy is applied to develop the image manifold at the different levels of semantic granularity i.e. two kinds of manifold (object manifold and scene manifold) are constructed using extended locally linear embedding and locally linear sub manifold extraction, considering the diversification of Web images [50]. The choice of specific rules to create graph connections and their weightings is crucial for the resulting segmentation hierarchy. A generic method to yield a hierarchical segmentation tree by iteratively applying minimum graph-cut to a weighted connectivity graph defined on a meaningful over-segmentation of RGB-D images. The approach allows for a task-dependent focus on the grouping hierarchy in order to identify task-relevant object parts, like handles or knobs for grasping and concavities for pouring [51]. A way to hierarchically group objects based on the Minimum Description Length principle. These groups convey higher order concepts which can be viewed as the building blocks of a scene. We show that using provide a significant increase (10%) in scene classification accuracy, thus proving that groups discovered for detections of these object groups as feature vectors used for scene classification can also be highly beneficial for computer vision tasks [52].

Hierarchical matching pursuit uses the matching pursuit encoder to build a feature hierarchy that consists of three modules: batch tree orthogonal matching pursuit, spatial pyramid matching, and contrast normalization. This hierarchical matching pursuit performs better than SIFT based single layer sparse coding and other hierarchical feature learning approaches: convolutional deep belief networks, convolutional neural networks and de-convolutional networks [53]. To incorporate the taxonomy information into deep learning framework, two deep neural network (DNN) based hierarchical learning methods for the acoustic scene classification task has been proposed. The first approach, hierarchical pre-training, a supervised learning process, can help the second DNN to get a better initialized weights based on the learning experience from the three high-level coarsely classified classes. The second approach, the multilevel objective function inspired by the multi-task learning [54].

A medium-size collection of geo-tagged photos, and a compact ontology of events and scenes for consumers to annotate photo collections instead of single image. A conditional random field based model that accounts for two types of correlations: (1) correlation by time and GPS tags and (2) correlation between scene- and event-level labels [55]. To extract a latent semantic structure of images for classification into categories using the singular value decomposition with banded color correlogram has been proposed. This correlogram is more suitable for the image classification task of color histogram [56]. To identify the top-down approach as an approach concerning mainly the testing phase of a classification algorithm, to a large extent independent of the particular local approach used for training. A tree-structured class hierarchies is preferred than Directed Acyclic Graph due to simpler structures. A type of hierarchical classification approach is better than other classification approach [57].

To analyzing high-level events in soccer video by combing low level feature analysis with high level semantic knowledge, a hierarchical framework has been presented. The sports domain semantic knowledge encoded in the hierarchical classification not only reduces the cost of processing data drastically, but also significantly increases the classifier accuracy. The hierarchical framework enables the use of simple features and organizes the set of features in a semantically meaningful way [58]. The outdoor scenes are considered to verify the scene classification using Hierarchical Space Tiling (HST) to learn the structure of hierarchical and reconfigurable scene models by quantizing the space of configurations [59].

VII APPLICATION OF HSC

The following application of hierarchical scene classification (HSC) is listed.

(a) To know the level of risk of danger in order to avoid collision on the road.

(b) In the classification of Objects, Images and scenes.

(c) In the recognition of Objects, Images and scenes.

There are many applications other than listed above. But, our focus is to know the level of risk of danger in order to avoid collision on the road.

VIII SCENARIO STUDY

According to Go & Carroll [45], a scenario is a description that contains (1) actors, (2) background information on the actors and assumptions about their environment, (3) goals or objectives, and sequences of actions and events". Also the usage of scenarios in any field is quite different, but the elements of a scenario are similar. The paper [3] defines the scenario as follow.

A scenario describes the temporal development between several scenes in a sequence of scenes. Every scenario starts with an initial scene. Actions & events as well as goals & values may be specified to characterize this temporal development in a scenario. Other than a scene, a scenario spans a certain amount of time."

IX CONCLUSION

The concepts of these scenario, situations and scenes have been studied in order to accomplish classification of situations, scenario and scenes from unpredictable infinite possibility of situation is basically needed for the autonomous driving vehicles.

REFERENCES

1. Jaiprakash Narain Dwivedi, http://hdl.handle.net/10228/00006323
2. Cambridge University Press 2017, http://dictionary.cambridge.org/dictionary /english/situation
3. Simon Ulbrich, Till Menzel, Andreas Reschka, Fabian Schuldt and Markus Maurer, Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving, ITSC, 982-988, 2015.
4. St'ephanie Lef'evre, Ruzena Bajcsy and Christian Laugier, Probabilistic Decision Making for Collision

Avoidance Systems: Postponing Decisions, IEEE/RSJ International Conference on Intelligent Robots and Systems, 2013, Tokyo, Japan. 2013.

5. R. Manduchi, A. Castano, A. Talukder and L. Matthies, Obstacle Detection and Terrain Classification for Autonomous Off-Road Navigation, Autonomous Robots, Volume 18, Issue1, pp 81-102, January 2005.
6. Andreas Kurz, An autonomous vehicle which learns basic skills and constructs maps for navigation, Robotics and Autonomous Systems, Volume 14, Issues 2-3, Pages 171-183, May 1995.
7. Christian P. Janssen and J. Leon Kenemans, Multitasking in Autonomous Vehicles: Ready to Go?, AutomotiveUI'15, Nottingham, UK ACM 978-1-4503-3736-6, September 1-3, 2015.
8. A. Hermann and S. Lutz, Situation based data distribution in a distributed environment model, Intelligent Vehicles Symposium, 2007 IEEE, 2007.
9. Andreas Hermann, S. Matzka, and J. Desel, Using a proactive sensor system in the distributed environment model, Intelligent Vehicles Symposium, 2008 IEEE, 2008.
10. Andreas Hermann and Jorg Desel, Driving Situation Analysis in Automotive Environment, Proceedings of the 2008 IEEE International Conference on Vehicular Electronics and Safety Columbus, OH, USA. September 22-24, 2008.
11. Jesse Levinson, Jake Askeland, Jan Becker, Jennifer Dolson, David Held, Soeren Kammel, J. Zico Kolter, Dirk Langer, Oliver Pink, Vaughan Pratt, Michael Sokolsky, Ganymed Stanek, David Stavens, Alex Teichman, Moritz Werling, and Sebastian Thrun, Towards Fully Autonomous Driving: Systems and Algorithms, Intelligent Vehicles Symposium (IV), 2011 IEEE, 05 July 2011.
12. Christian Berger and Bernhard Rumpe, Autonomous Driving-5 Years after the Urban Challenge: The Anticipatory Vehicle as a Cyber-Physical System, proceedings of the 27th IEEE/ACM International Conference on Automated Software Engineering (ASE 2012) : Essen, Germany, 3 - 7 September 2012.
13. Simon Ulbrich and Markus Maurer, Situation Assessment in Tactical Lane Change Behavior Planning for Automated Vehicles, ITSC, 975-981, 2015.
14. Dean A. Pomerleau, Efficient Training of Artificial Neural Networks for Autonomous Navigation, Massachusetts Institute of Technology, Spring 1991, Vol. 3, No. 1, Pages: 88-97 Posted Online March 13, 2008.
15. Aude Olivia, Antonio Torralba, Modeling the shape of the scene: a holistic representation of the spatial envelope, Int. J. Comput. Vision 42(3), 145–175, 2001.
16. B V V Sri Raj Dutt, Pulkit Agrawal and Sushoban Nayak, Scene Classification in Images, https://people.eecs.berkeley.edu/~pulkitag/scene_report.pdf.
17. Li Fei-Fei and Pietro Perona, A bayesian hierarchical model for learning natural scene categories, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 2, pp. 524–531, IEEE, 2005.
18. Li-Jia Li and Li Fei-Fei, What, where and who? Classifying events by scene and object recognition, 11th International Conference on Computer Vision, ICCV 2007, pp. 1–8, IEEE, 2007.
19. Svetlana Lazebnik, Cordelia Schmid, Jean Ponce, Beyond bags of features: spatial pyramid matching

for recognizing natural scene categories, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2169–2178, IEEE, 2006.

20. Derek Hoiem, Alexei A. Efros, Martial Hebert, Geometric context from a single image, Tenth IEEE International Conference on Computer Vision, ICCV 2005, vol. 1, pp. 654–661, IEEE, 2005.
21. Antonio Torralba, Rob Fergus and William T. Freeman, 80 million tiny images: a large dataset for non-parametric object and scene recognition, IEEE Trans. Pattern Anal. Mach. Intell. 30(11), 1958–1970, 2008.
22. George A. Miller, WordNet: A Lexical Database for English, ACM 38(11), 39–41, 1995.
23. Julia Vogel and Bernt Schiele, Natural scene retrieval based on a semantic modeling step, In Conference on Image and Video Retrieval CIVR 2004, Dublin, Ireland, July 2004.
24. Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn and Andrew Zisserman, The PASCAL Visual Object Classes (VOC) Challenge, Int. J. Computer Vision 88(2), 303–338, 2010.
25. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li and Li Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Database, IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 248–255, IEEE, 2009.
26. Bryan C. Russell, Antonio Torralba, Kevin P. Murphy, William T. Freeman, LabelMe: a database and web-based tool for image annotation, Int. J. Comput. Vision 77(1–3), 157–173, 2008.
27. Jianxiong Xiao, James Hays, Krista A. Ehinger, Aude Oliva and Antonio Torralba, SUN Database: Large-scale Scene Recognition from Abbey to Zoo, IEEE conference on Computer Vision and Pattern Recognition (CVPR), pp. 3485–3492, IEEE, 2010.
28. Greg Griffin, Alex Holub and Pietro Perona, Caltech-256 object category dataset, Technical Report CNS-TR-2007-001, California Institute of Technology, 2007.
29. Bolei Zhou, Agata Lapedriza1, Jianxiong Xiao, Antonio Torralba, and Aude Oliva, Learning Deep Features for Scene Recognition using Places Database, Advances in Neural Information Processing Systems, pp. 487–495, 2014.
30. C. Chen et al., Chapter 2, Scene Understanding Datasets Big Visual Data Analysis, SpringerBriefs in Signal Processing, 2016.
31. Zhen Zuo, Gang Wang, Bing Shuai, Lifan Zhao, Qingxiong Yang and Xudong Jiang, Learning Discriminative and Shareable Features for Scene Classification, European Conference on Computer Vision, 2014.
32. Anna Bosch, Andrew Zisserman and Xavier Munoz, Scene Classification via PLSA, European Conference on Computer Vision, ECCV 2006.
33. Haohui Yin, Scene classification using spatial pyramid matching and hierarchical Dirichlet processes, Thesis, Rochester Institute of Technology, 2010.
34. Rong Yan, Yan Liu, Rong Jin and Alexander Hauptmann, On Predicting Rare Class with SVM Ensemble in Scene Classification, International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003), Hong Kong, China, April 6-10, 2003.
35. Jiebo Luo and Matthew Boutell, Natural scene classification using overcomplete ICA, Pattern

Recognition Society, Elsevier Ltd., 38 (2005) 1507 – 1519, 2005.

36. Matthew Boutell, Xipeng Shen, Jiebo Luo and Chris Brown, Multi-label Semantic Scene Classification, technical report, dept. comp. sci. u. Rochester, 2003.
37. Selim Aksoy, Krzysztof Koperski, Carsten Tusk, Giovanni Marchisio and James C. Tilton, Learning Bayesian Classifiers for Scene Classification with a Visual Grammar, IEEE Transactions on Geoscience and Remote Sensing, Vol. 43, Issue 3, Pages 581-589, Mar 2005.
38. Xin Li and Yuhong Guo, Latent Semantic Representation Learning for Scene Classification, Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 2014.
39. Jincheng Huang, Zhu Liu and Yao Wang, Joint Scene Classification and Segmentation Based on Hidden Markov Model, IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 7, NO. 3, JUNE 2005.
40. Jun Yang, Yu-Gang Jiang, Alexander Hauptmann and Chong-Wah Ngo, Evaluating Bag-of-Visual-Words Representations in Scene Classification, Proceedings of the international Workshop on Workshop on Multimedia information Retrieval , 197-206, 2007.
41. Jian Yao, Sanja Fidler and Raquel Urtasun, Describing the Scene as a Whole: Joint Object Detection, Scene Classification and Semantic Segmentation, Conference of Computer Vision and Pattern Recognition (CVPR), 2012.
42. Mayank Juneja, Andrea Vedaldi, C. V. Jawahar and Andrew Zisserman, Blocks that Shout: Distinctive Parts for Scene Classification, Computer Vision and Pattern Recognition (CVPR), IEEE Conference, 2013.
43. Cambridge University Press 2017, http://dictionary.cambridge.org/dictionary/english /scenario
44. K. Go and J. M. Carroll, The blind men and the elephant: Views of scenario-based system design, interactions, vol. 11, no. 6, pp. 44-53, 2004.
45. Cambridge University Press 2017, https://dictionary.cambridge.org/dictionary/english/scene
46. Oden Maron and Aparna Lakshmi Ratan, Multiple-Instance Learning for natural scene classi_cation, In The Fifteenth International Conference on Machine Learning, pp 341-349, 1998.
47. S. Battiato, G. M. Farinella, G. Gallo and D. Ravi, Spatial Hierarchy of Textons Distributions for Scene Classification, International Conference on Multimedia Modeling, pp 333-343, 2009.
48. Clement Farabet, Camille Couprie, Laurent Najman and Yann LeCun, Learning Hierarchical Features for Scene Labeling, IEEE Transactions on Pattern Analysis and Machine Intelligence, in press, 2013.
49. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE, 86(11):2278–2324, November 1998.
50. Rong ZHU, Min YAO, Li-hua YE and Jun-ying XUAN, Learning a hierarchical image manifold for Web image classification, Journal of Zhejiang University-SCIENCE C (Computers & Electronics), 2012.
51. Andre Uckermann, Christof Elbrechter, Robert Haschke and Helge Ritter, Hierarchical Scene Segmentation and Classification, International Conference on Intelligent Robots and Systems, Chicago,

USA, IROS 2014.

52. Amir Sadovnik and Tsuhan Chen, HIERARCHICAL OBJECT GROUPS FOR SCENE CLASSIFICATION, IEEE International Conference on Image Processing, ICIP, 2012.
53. Liefeng Bo, Xiaofeng Ren and Dieter Fox, Hierarchical Matching Pursuit for Image Classification: Architecture and Fast Algorithms, Advances in Neural Information Processing Systems (NIPS), December, 2011.
54. Yong Xu, Qiang Huang, Wenwu Wang and Mark D. Plumbley, HIERARCHICAL LEARNING FOR DNN-BASED ACOUSTIC SCENE CLASSIFICATION, Detection and Classification of Acoustic Scenes and Events, 2016.
55. Liangliang Cao, Jiebo Luo, Henry Kautz and Thomas S. Huang, Annotating Collections of Photos Using Hierarchical Event and Scene Models, CVPR, pages 1-8, 2008.
56. Jing Huang, S Ravi Kumar and y Ramin Zabih, An Automatic Hierarchical Image Classification Scheme, EURASIP Journal on Applied Signal Processing, 2003.
57. Carlos N. Silla Jr and Alex A. Freitas, A Survey of Hierarchical Classification Across Different Application Domains, Data Mining and Knowledge Discovery, Volume 22, Issue 1, pp 31–72, January 2011.
58. M. H. Kolekar and K. Palaniappan, A Hierarchical Framework for Semantic Scene Classification in Soccer Sports Video, TENCON, IEEE Region 10 Conference, 1-6, 2008.
59. Shuo Wang, Yizhou Wang and Song-Chun Zhu, Hierarchical Space Tiling for Scene Modeling, Volume 7725 of the series Lecture Notes in Computer Science, pp 796-810, Computer Vision – ACCV 2012.