Sri Lanka Institute of Information Technology

Image content based classification of Vacation/ tourism related images

Project ID: 16-074

Project Proposal

B.Sc. Special (Honors) Degree in Information Technology

Submitted on 10/03/2016

Image content based classification of vacation/ tourism related images

Project ID :16-074

Authors:

Student ID	Name	Signature
IT13047566	Amjad M.N.M.	
IT13018238	Gunathilaka M.A.S.S	
IT13081744	Weligampola L.P.	
IT13022570	Gunasinghe S.U.	

Supervisor

……

Dr. Rohana Thilakumara

DECLARATION

We declare that this is our own work and this project proposal does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

……..
…….

Amjad M.N.M

Gunathilaka M.A.S.S

……..
…….

Weligampola L.P.

Gunasinghe S.U.

ABSTRACT

Grouping images into meaningful categories using low-level visual features is a challenging and important problem in content-based image retrieval and classification. Humans look at images and easily to identify the objects and scenarios in the images, using the knowledge and experience of what they have seen and gained thorough their lifetime but computer programs cannot clearly identify what the images represents and what the contents of the images are. This is the main problem we are trying to solve in this research project. The proposed program will segment a given image/s (related to travel/tourism) and from the segments create objects then classify the objects, and from the objects that relate to each other, identify the context of the given image/s and classify them accordingly. The program will be able to output a classified set of images or when given an image, describe the given image, specify what kind of image it is, indoor/outdoor, nature, city etc.

TABLE OF CONTENT

Contents

1. INTRODUCTION

1.1 Background

As humans it is easy for us to recognize the contents of a picture when we are shown one and classify them broadly into a general category of images they belong to, using our knowledge and past experience. For example when given a set of images, we could easily identify which ones are beaches, which ones are pictures of cities, which ones are pictures of sunsets etc. For human beings this task is not that big a deal and we don't think twice before understanding and interpreting an image. But it is a very difficult task for a computer program to do.

Understanding the contents of an image, interpreting the image and grouping them into semantically meaningful categories by understanding the correlation between the objects in the images and using low-level visual features is a challenging and important problem in content-based image understanding and classification using computer programmers. Even though there are lots of algorithms available to identify the main object or objects in an image, there is a lack of algorithm to identify the context of the image or the image as a whole, i.e.; whether it is a picture of a waterfall or a picture of a sunset etc.…

With the development of social networks and digital photography more and more people are storing their photographs of places they have visited and captured. As an example if we consider a travel agency, they are interested in photographs of various sites, that user can search and plan where they want to go. These images will be more useful if they are automatically categorized and classified this was the basis for us to develop an algorithm for categorizing the huge amount of image data or a website that specializes in providing users with travel destinations according to the user's preferences, this site could use the data obtained from the users images (for example from Facebook) and use that data to provide semantically accurate travel routes and destinations.

An application that could interpret an image and identify the scenery according to the context of the objects in the image would be of great value to all of us.

1.2 Literature Review

Content based image identification and classification can be done in several ways such as using Random Forest/ferns, Neural Networks, Artificial intelligence, Relevance Vector Machine (Machine learning technique that uses Bayesian inference) or a combination of these methods. Several researchers have introduced advanced algorithms to classify images based on their uses but most of them have limitations and issues of their own, or have since run out of date with the changing technological landscape.

Image segmentation is the process of partitioning a digital image into multiple segments on a spatial plain or process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics. Even though some researchers have developed algorithms to segment the image, there is still a problem of simultaneously segmenting an image into its constituent semantic regions which is identified by F. Schroff , A.Criminisi , A.Zisserman. They used random forest as the algorithm for image segmentation and improved it in order to overcome that problem. [1] are trying to identify an object in an image by automatically selecting the Region of Interest (ROI),the most important part of the image, or the main object in the image by selecting a rectangular area of the image as a Region of Interest (ROI). The idea is that between a subset of the training images for a particular class there will be regions with high visual similarity (the object instances). These regions can be identified from the clutter by measuring similarity using the spatial pyramid representation as stated in [2] S. Lazebnik, C. Schmid, and J. Ponce's research paper titled "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories"

The main problem in [1] is that they only find the main object in the image or the ROI and not the sub objects in the image. For our application to work we would have to modify the algorithm or use another appropriate algorithm to identify all or most of the significant object in a given image. As to identify the relationships between those objects and infer the context of the image from them.

The Second major part of image classification is identifying the objects in the image from the segmented image data. For this the image should be represented as both appearance and shape according to [1] the hard part is that images can be identified by appearance or shape or both.

The classifier should be able to identify images from their appearance or shape. For example, a car can be better distinguished by its shape and a lion by its appearance [1].

As for the classifier [1] have used random forest classifier for object recognition and improved the algorithm to work for a broader classes of objects. [1] the advantage of randomized trees is that they are much faster in training and testing than traditional classifiers.They also enable different cues(such as appearance and shape) to be "effortlessly combined" [5]. [1] have used the datasets namely caltech-101 and caltech-256 which contain most of the training images needed to create their image classifier algorithm.

In addition to random forest [1] also use random ferns classifier by Ozuysal in [6] to increase the speed of random forest classifier.[1] explains that Ferns are non-hierarchical structures where each one consists of a set of binary tests. During training there are an ordered set of tests applied to the whole training data set. This is in contrast to random forests where only the data that falls in a child is taken into account in the test. As in random forests "leaves" store the posterior probabilities. During testing the probability that an image belongs to any one of the classes that have been learned during training is returned. The result of each test and the ordering on the set defines a binary code for accessing the "leaf" node. As in random forests, the test image is passed down all the randomized ferns. Each node in the fern provides a result for the binary test which is used to access the leaf which contains the posterior probability. The posteriors are combined over the ferns in the same way as for random forests over trees.

On the subject of multiple object recognition from an image we came across an interesting research project [7] done at google recently, which has managed to improve the problem of multiple object recognition effectively, they have made use of customized convolutional neural network. For the detection part of the application, an improved neural network model was used in the sophisticated R-CNN detector by Ross Girshick et al [9], with additional help from the multibox method [8]. For the classification part,the research; several ideas from the work of Andrew Howard [10] were incorporated and extended, specifically as they relate to image sampling during training and evaluation.

Another interesting research on image classification using deep convolutional neural network (CNN) [11] they have trained a large deep CNN to classify 1.2 million high resolution images. Results obtained in [11] show that a large, deep CNN is capable of achieving record-breaking results on a highly challenging dataset using purely supervised learning. They have noted that

their network's performance degrades if a single convolutional layer is removed. For example they state, removing any of the middle layers results in a loss of around 2% for the top 1 performance of the network. So the depth really is important for achieving results according to [11]

According to the the research paper on Detection and Classification of Multiple Objects using an RGB-D Sensor and Linear Spatial Pyramid Matching [12] ,they have used an RGB-D sensor such as the Microsoft Kinect sensor [12] to capture an image with depth information too and used that image to generate a depth map to identify objects and background from the image. As the first step they capture a depth image image [12] then The normalized depth image is processed with convolution by a two dimensional Gaussian filter to produce the gradient image where edges of the objects have very large values and all other area having very small values.[12] Then the objects are identified in stages with defined parameters [12]. For the image classification parts they have use Linear Spatial Pyramid Matching (LSPM)

We chose to mention these Research papers and documents here since they contain elements we will need to study more and incorporate into our own research

1.3 Research Gap & Research Problem

Grouping images into (semantically) significant categories using low-level visual features is a challenging and important problem in content-based image retrieval and classification. Humans look at images and easily identify the objects and sceneries in the images, using the knowledge and experience of what they have seen and gained through their lifetime and categorize or group those images into high level categories. But computers or more specifically computer programs cannot clearly identify what the images represents and what the contents of the images are. Even though algorithms have been developed to identify objects in an images, and even some to identify multiple objects in an image to an extent they are not capable of interpreting the inter relation between the objects in the images and thereby the overall context of the image, or in other words what the picture is about as an overview. This is a major requirement in intelligent web applications that need to classify and if possible identify the content in images quickly and reliably.

We are researching to provide an intelligent image processing application that is able to identify objects in an image and scenarios in the images using the objects in the images and classify the objects in the images and Group the images into (semantically) meaningful categories using low-level and high-level visual features. Specifically, we consider the hierarchical classification of vacation images; at the highest level, images are classified as indoor or outdoor; outdoor images are further classified as city,village,town or landscape etc; finally, a subset of landscape images is classified into waterfalls,beaches, sunsets/sunrise, forest, and mountain etc classes. The application will be able to identify the objects and features in those images and classify them accordingly to be used by other web applications. Furthermore the application will be able to learn(train the system) from the images it classifies and will be able to quickly and accurately identify the images as it progresses. The algorithm will be demonstrated for identifying tourist attraction locations in images and classifying them.

2. OBJECTIVES

2.1 Main Objectives

- The user should be able to upload an image or a set of unclassified images related to traveling/tourism such as beaches, sunsets, waterfalls, landscapes, mountains etc.. and the application should output a classified list of images accurately grouping each image into broader category they belong to.
- The user should be able to get a list of objects identified from the image ,a detailed description of the image and any other assumptions the program can come to from the image in addition to the main classifying classes. these information can be used for future semantic web purposes

2.2 Specific Objectives

- Create or improve an existing appropriately customized algorithm for image segmentation which is the process of partitioning a digital image into multiple segments on a spatial plain.
- Create or improve an existing appropriately customized algorithm for object formation from previously segmented images parts. The segmented image parts are organized and objects are identified from them.
- Create or improve an existing appropriately algorithm for image recognition and identification. The pre formed objects are identified and labeled.
- Identify the correlation between the objects previously identified from the images.
- Create or improve an existing appropriately for context identification and classification. The identified objects and images are analyzed in relation to each other and the context of the image is identified and classified.

3. RESEARCH METHODOLOGY

In this section we will discuss the methodology we will be using to carry out the research project and for managing the project. We will try our best to present a systematic approach to the research, project development and implementation in order to create a successful solution within the allocated time, budget, and using currently available tools and technologies.

3.1 System Diagram

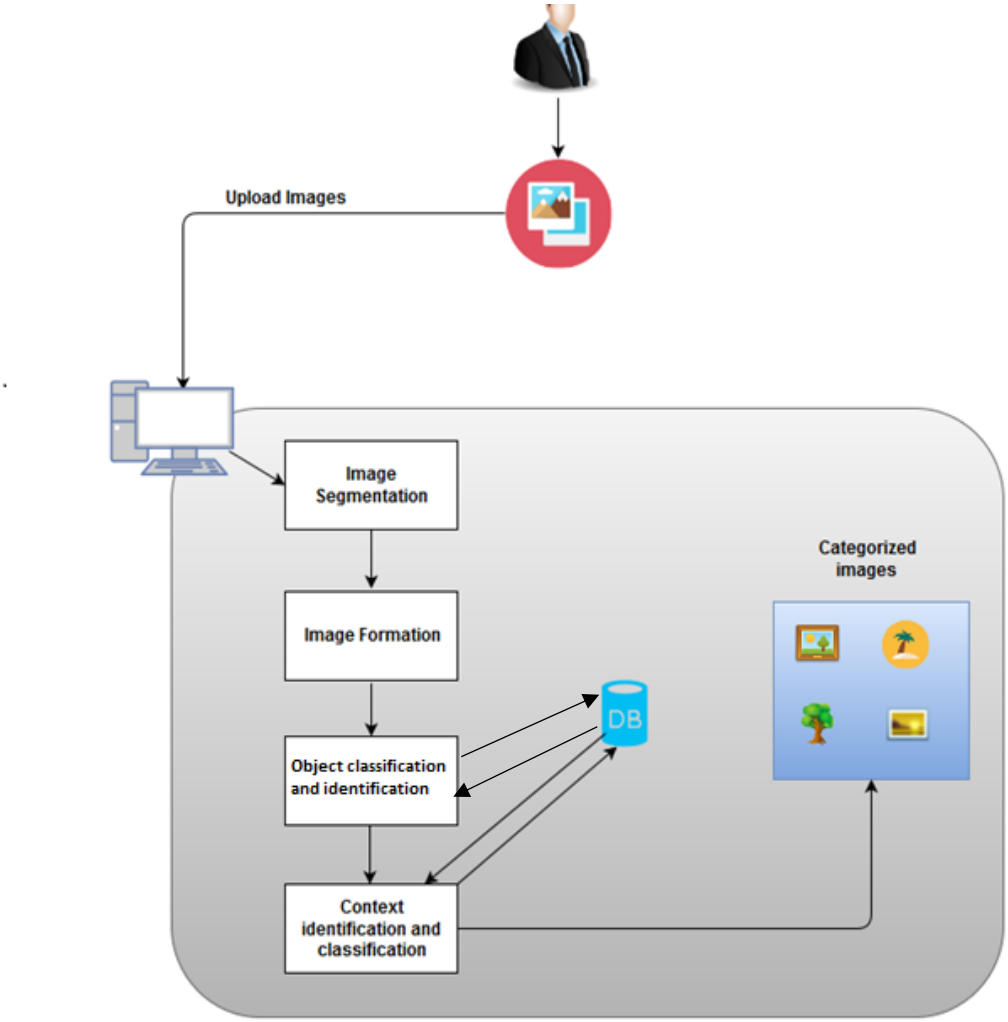

Figure 3.1.1 System Diagram

3.2 Software Development Life Cycle

We are going to use “Iterative waterfall model” as the development cycle method because of having a proper mechanism to handle the requirement changes.

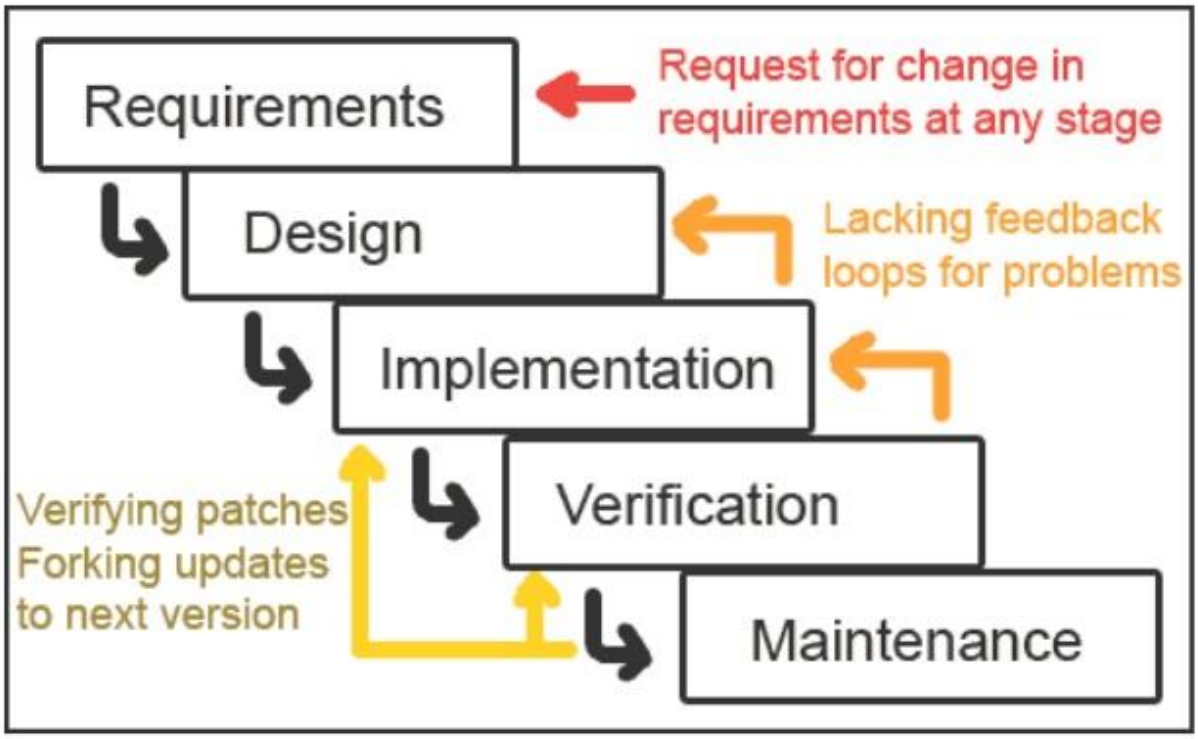

Figure 3.2.1 Software development life cycle

3.3 System Description

The proposed system will be able to classify input images into meaningful categories after processing the input images. You can provide large number of images to the system and get them organized into groups of related images and a detailed description of the image.. We can direct the processed result to any application where you need to find the category such as sunset, beaches, mountains, forests etc. This can be useful to create a semantic web experience for users when they visit websites and such.

3.3.1 Image spatial representation and segmentation

Image pre-processing

The aim of image pre-processing is to suppress undesired distortions or enhance some image features that are important for further processing or analysis. Some noise will appear on the image. Image noise is the random variation of brightness in images. Removing the noise is an important step when image processing is being performed [13].

Figure 3.3.1.1 Pre-processed image

Image segmentation

Image segmentation is the process of dividing an image into multiple parts or multiple regions. It is the first step in image identification and classification. There are two majorly used techniques in image segmentation which are namely edge based segmentation and region based segmentation. There are several methods in edge based segmentation such as thresholding

methods, k-means clustering, histogram based methods etc. Most of the researches have used edge based segmentation technique because of the simplicity of it.

Toru Tamaki, Tsuyoshi Yamamura and Noboru Ohnishi have used "split-and-merge" method for image segmentation which overcome the problem stated in [4]. Histogram based method is another method which is very efficient compared with other methods. According to Linda G. Shapiro and George C. Stockman, A histogram is computed from all of the pixels in the image, and the peaks and valleys in the histogram are used to locate the clusters in the image[3]. Even though above method is efficient, there is a difficulty of identify significant peaks and valleys in the image.

Another method used for image segmentation is the "split-and-merge" method. Split and merge method has been around for a while now. It has a pyramidal structure which is shown in figure 3.3.1.2.

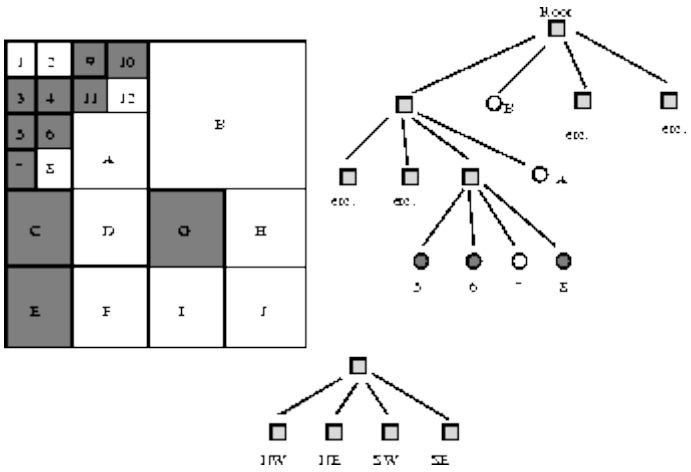

Figure 3.3.1.2 Pyramidal structure of splitting and merge technique

Another algorithm used for image segmentation is the popular random forest algorithm. A random forest multi-way classifier consists of a number of trees, with each tree grown using some form of randomization. The leaf nodes of each tree are labeled by estimates of the posterior distribution over the image classes. Each internal node contains a test that best splits the space of data to be classified. An image is segmented by sending each pixel down every tree and aggregating the reached leaf distributions.[1] using this method the pixels can be grouped into super pixels and segments of the image can be formed.

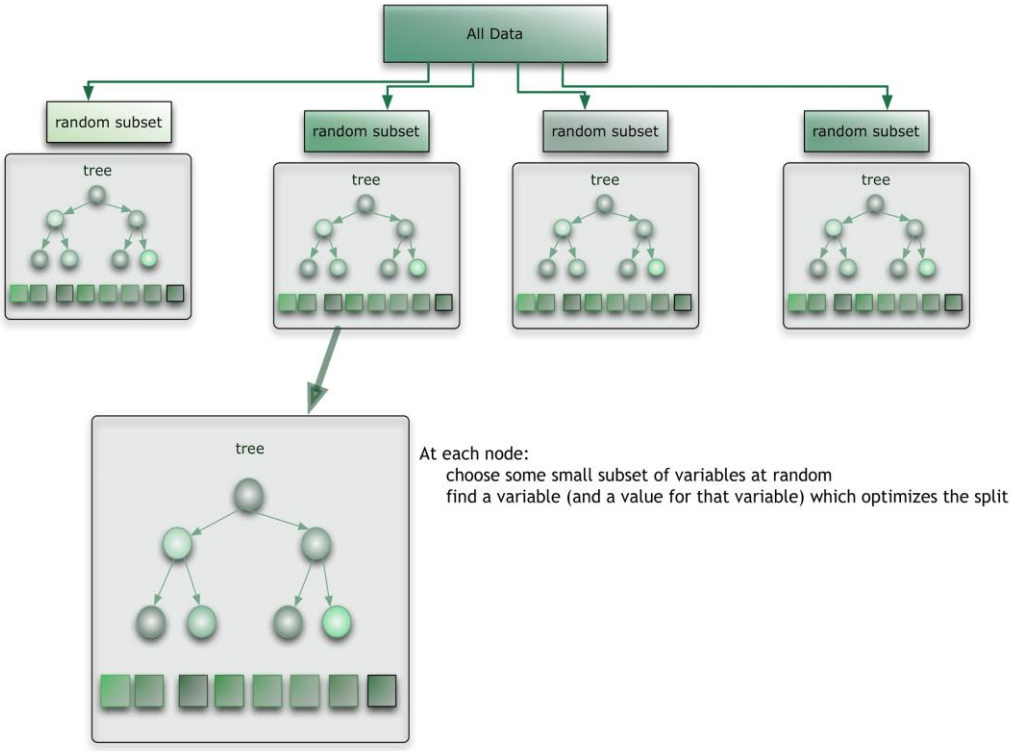

Figure 3.3.1.3 random forest classification

Below figure 2.2.1.3 represents the source image which needs to be segmented. The other image represents the segmented image of the above image. The second image clearly shows the frames separately.

Figure 3.3.1.4 original image (left) and segmented image (right)

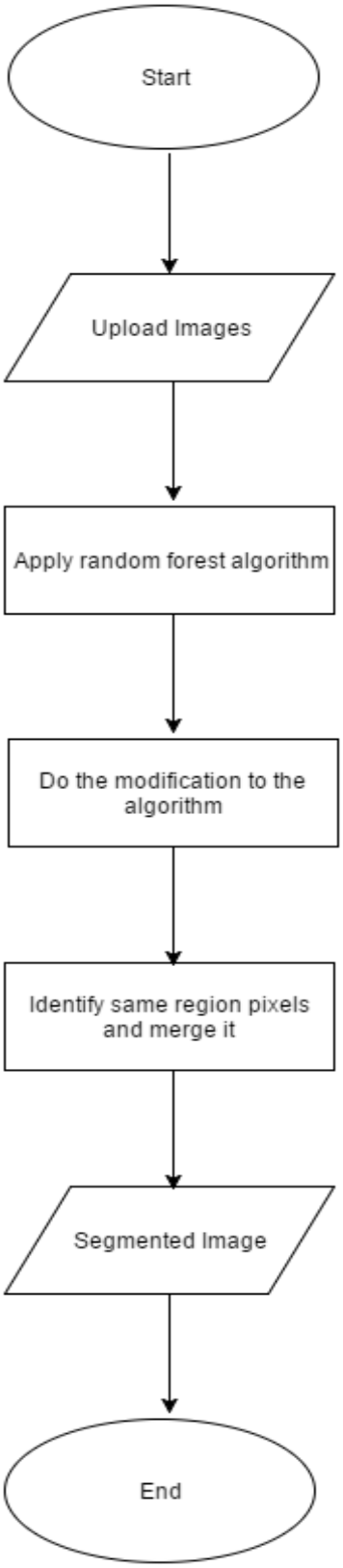

Figure 3.3.1.5 segmentation flow chart

3.3.2 Object formation from segmented images parts

Object formation is used after segmentation to identify the similar regions and their surrounding regions and group them in order to identify objects clearly. Previously segmented image has set of different segmented layers which are separated based on color, regions, curves etc… In here we need to form the segmented parts into meaningful objects.

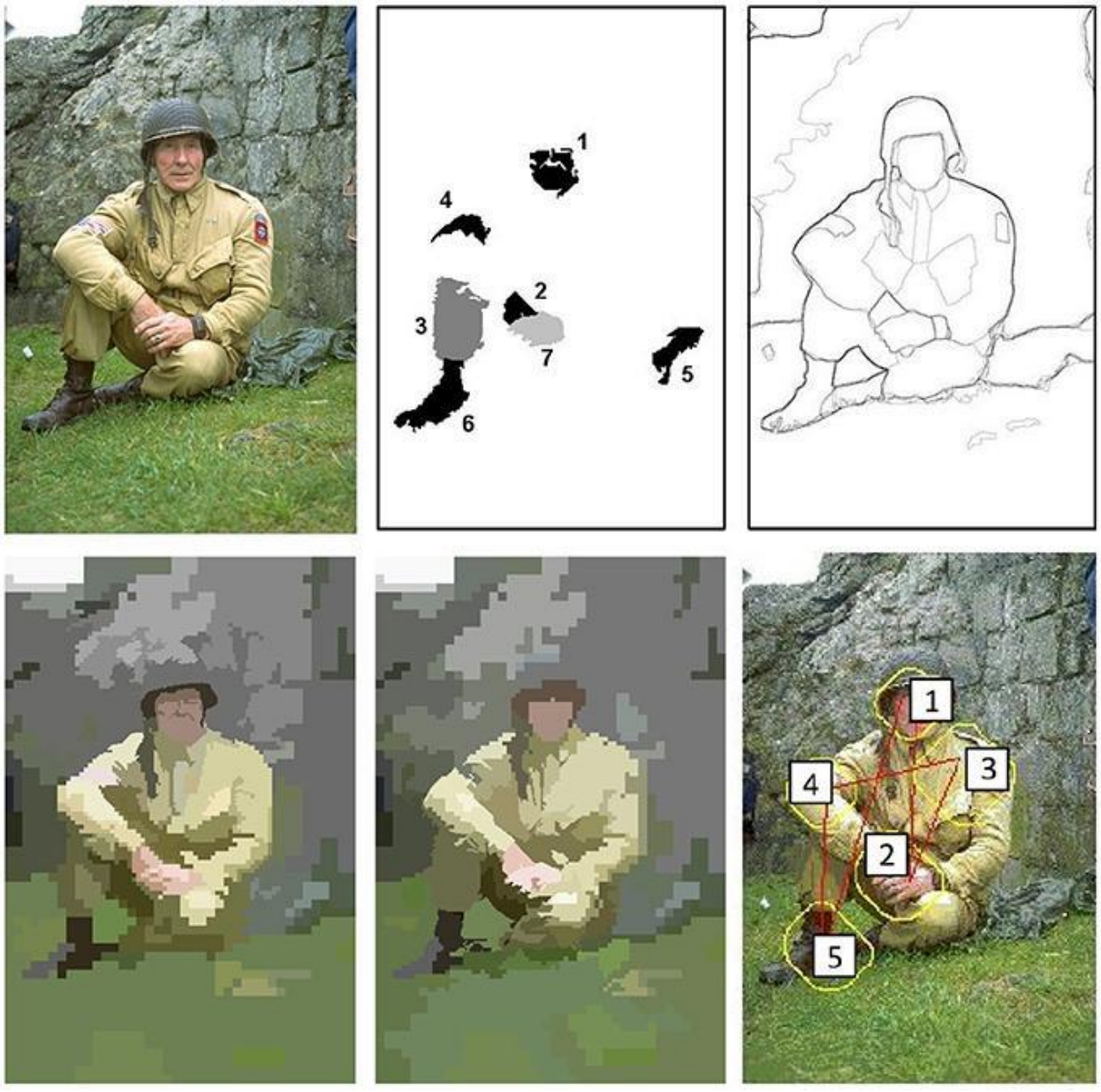

Figure 3.3.2.1 Image formation

Figure 3.3.2.1 shows a segmented image which has several segmented layers. But the problem is that we cannot identify objects by just taking layers separately. We need a proper

mechanism for identifying correlation between these layers and group them in order to come up with a meaningful object. For this purpose we need image formation process.

Most of the time it is a difficult task to form the layers because of having large no of segmented layers. Because of this problem [4] has used geometrical features like smoothness, area ratio, inclusion and continuity. So that if one of these features is satisfied the condition, they are merged together. The segmented layers are taken one by one and check the geometrical features like smoothness, ratio to identify the similar segmented layers in the same object.

Furthermore Richard Socher in his thesis [15] has used a neural network to map the image segments into "semantic" space. And using these semantic region representations as input, he has used his own RNN algorithm to merge semantically related neighboring regions to form a recognizable object as shown in figure 3.3.2.2 below.

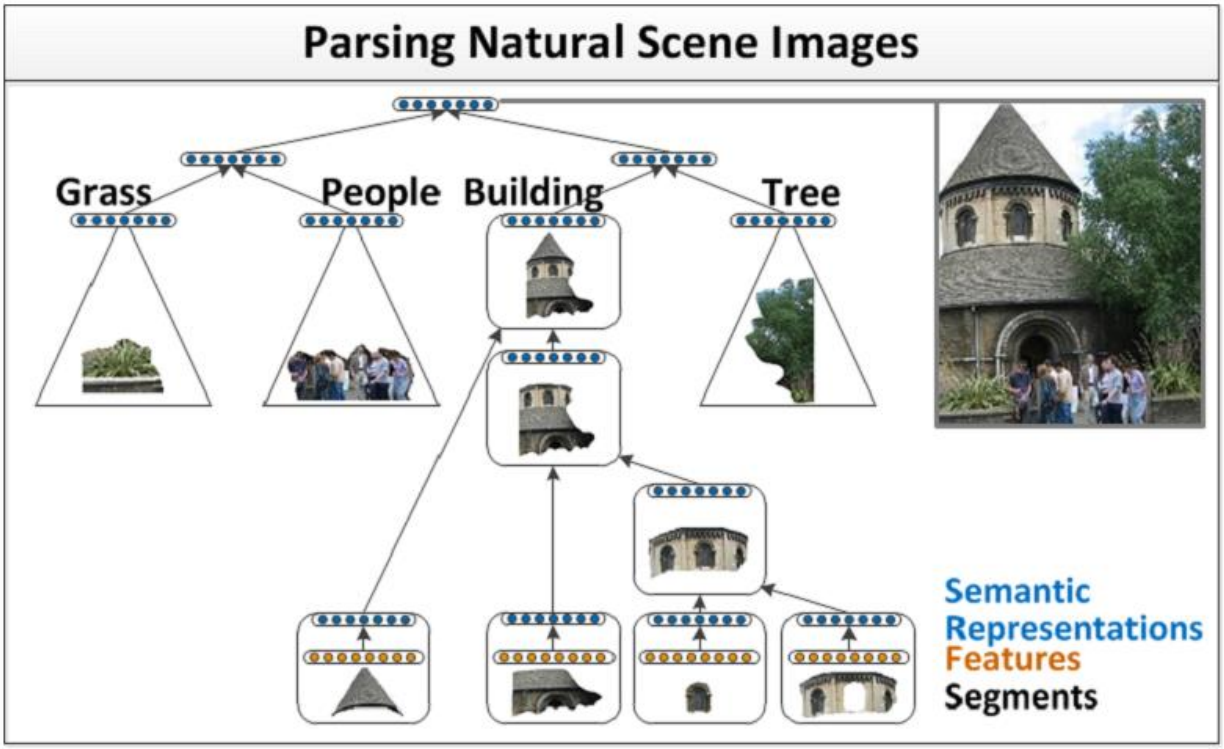

Figure 3.3.2.2 Forming an image with segmented parts[15]

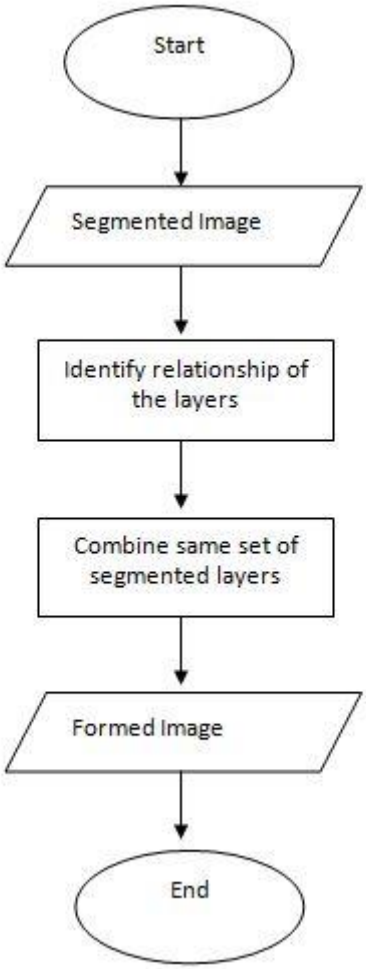

Figure 3.3.2.3 object formation flow chart

3.3.3 Object classification and identification

Object identification and recognition is another important part in image classification which is used to identifying objects in an image. Basically there are two pattern recognition algorithms called appearance-based and feature based. There are many features which are interesting points on the object that can be extracted to provide a feature description of the object. This description extracted from training images can be used to identify the object when attempting to identify an object in an image containing many other objects.

Machine learning

Machine learning is the study of computer algorithms that improve automatically through experience and has been central to AI research since the field's inception.

Unsupervised learning is the ability to find patterns in a stream of input. Supervised learning includes both classification and numerical regression. Classification is used to determine what category something belongs in, after seeing a number of examples of things from several categories. Regression is the attempt to produce a function that describes the relationship between inputs and outputs and predicts how the outputs should change as the inputs change. In reinforcement learning the agent is rewarded for good responses and punished for bad ones.

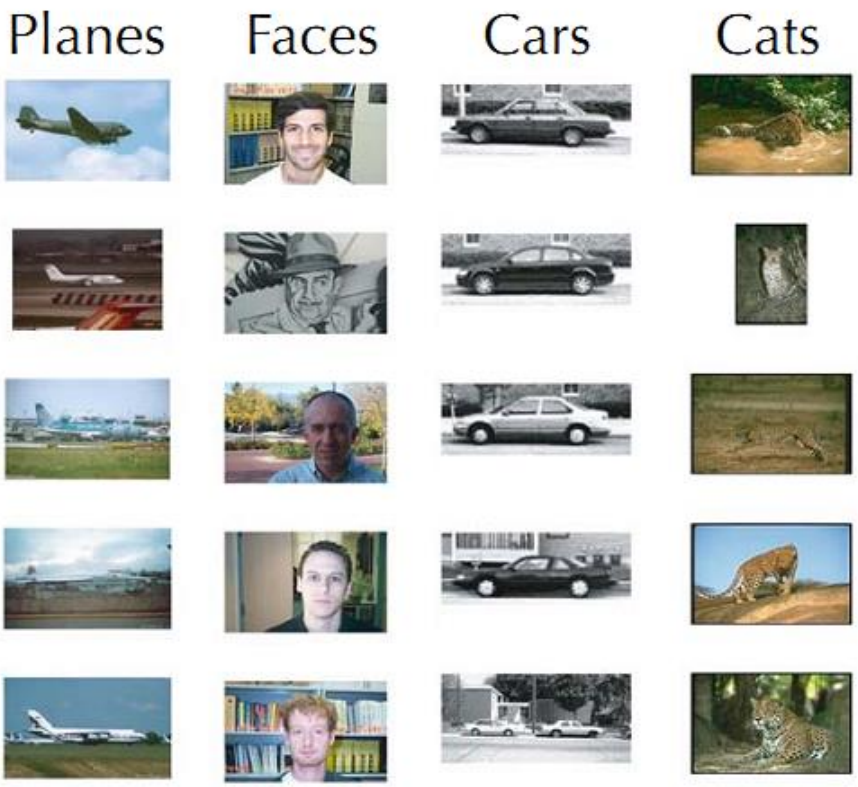

Figure 3.3.3.1 sample images used to train the system

For the purpose of object classification we are planning to use machine learning techniques like Recursive Neural Network for training [15] or Convolutional Neural Networks as used in[10][11] A set of images divided into classed and labeled appropriately can be used to train the network and after successfully training the system, the system will be able to identify objects in input images obtained from the previous phase.

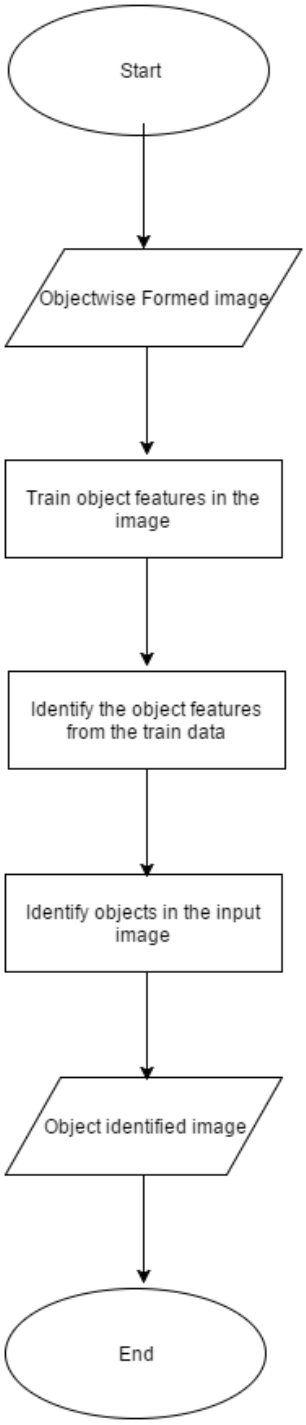

Figure 3.3.3.2 image identification and classification flow chart

3.3.4 Image Context identification and classification

Context identification and classification is the most important part in this our proposed research as it is needed to provide the user with the classification category/s for which the input image/s belongs to. Researches have been conducted and are still being conducted on image

classification based on high and low level analysis of image content (colour, texture, objects). Our research focuses on identifying the objects in the image and along with image content information and image classification properties, to improve image classification and identify the interrelationships between the objects in the images and their overall contextual category and provide a holistic understanding of the image.

Artificial intelligence and neural network methods/algorithms can be used to identify the correlation between objects identified in the previous phase using random forest classifier. The objects identified in the previous phase can be used as the input in this phase and the Neural network model can be trained to recognize the inter-relation between the objects. Furthermore an artificial intelligence component can also be used to further clarify and understand the objects in the image and get a holistic idea of the image.

Use of more training images will increase the accuracy of classification. We can classify the sample images into categories which will be stored in the database for classification. They will be compared against the user input.

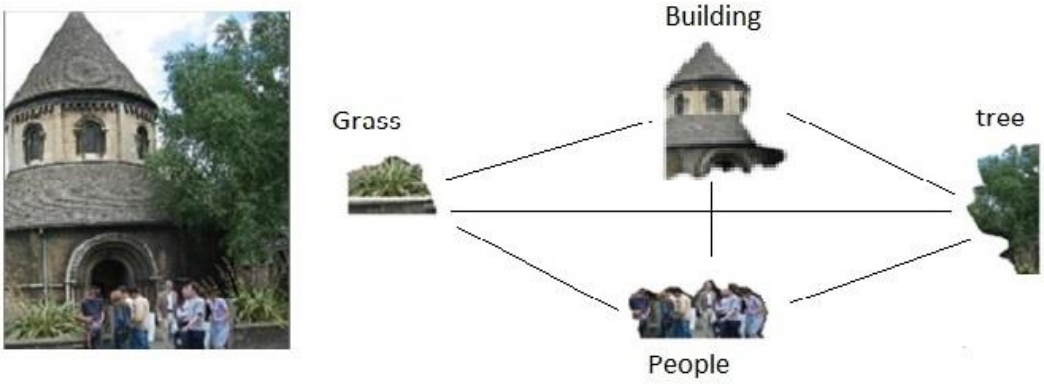

Figure 3.3.4.1 object content classification

3.4Gantt chart

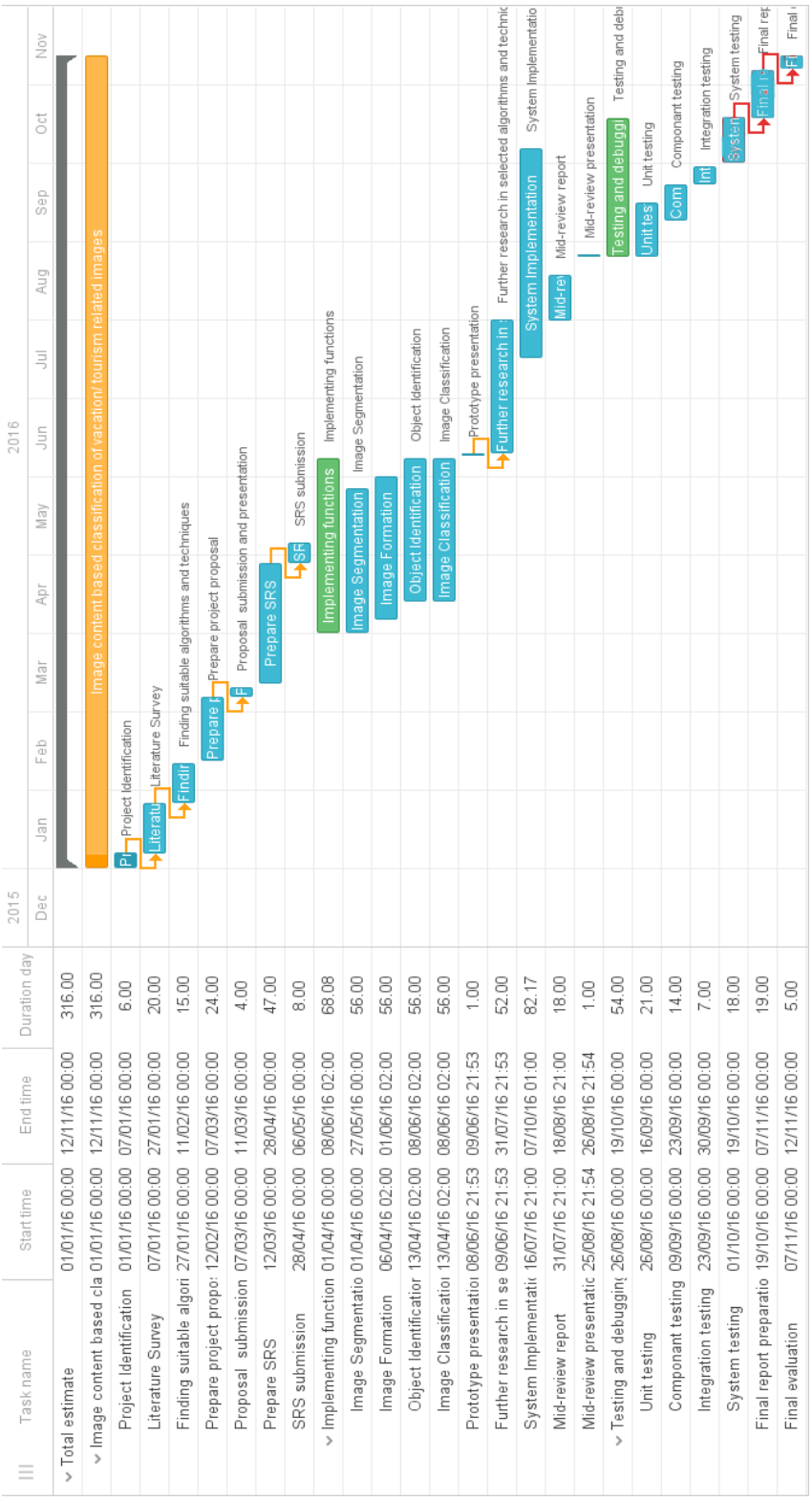

Figure 3.5.1 Gantt chart

4. DESCRIPTION OF PERSONAL AND FACILITIES

Member	Component	● Task
IT13047566 - Amjad M.N.M.	Image Context identification and classification	● identify relationships between objects using machine learning techniques ● Deduce the category the image as a whole belongs to using Artificial intelligence
IT13018238 - Gunathilaka M.A.S.S	Object classification and identification	● Use random forest algorithm for train the images ● Identify the patterns and match objects with the user input image.
IT13081744 - Weligampola L.P.	Object formation from segmented image parts	● Segmented parts are reorganized to create objects. ● Reverse neural networks are used to semantically relate neighbouring regions.
IT13022570 - Gunasinghe S.U.	Image segmentation	● Split the image into small pixels(super pixels) ● Merge the same region pixels based on colors, curves etc...

5. BUDGET

Details	Price (Rs.)
Stationary	2000.00
Traveling Cost	1000.00
Total Budget	3000.00

6. REFERENCES

[1] A. Bosch, A. Zisserman, and X. Munoz. Image Classification using Random Forests and Ferns

[2] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. CVPR, 2006.

[3] Linda G. Shapiro and George C. Stockman (2001): “Computer Vision”, pp 279-325, New Jersey, Prentice-Hall, ISBN 0-13-030796-3

[4] Toru Tamaki, Tsuyoshi Yamamura and Noboru Ohnishi: “Image segmentation and object extraction based on geometric features of regions”,Part of IS&T/SPIE Conf. on VCIP'99, SPIE Vol.3653, Part Two, pp.937{945 (1999 1)

[5] J. Winn and A. Criminisi. Object class recognition at a glance. CVPR, 2006.

[6] M. Ozuysal, P. Fua, and V. Lepetit. Fast keypoint recognition in ten lines of code. CVPR, 2007

[7] Research Blog, "Building a deeper understanding of images", 2016. [Online]. Available: http://googleresearch.blogspot.com/2014/09/building-deeper-understanding-of-images.html [Accessed: 28- Feb- 2016].

[8] Erhan D., Szegedy C., Toshev, A., and Anguelov, D., "Scalable Object Detection using Deep Neural Networks", *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2147-2154.*

[9] Girshick, R., Donahue, J., Darrell, T., & Malik, J., "Rich feature hierarchies for accurate object detection and semantic segmentation", *arXiv preprint arXiv:1311.2524, 2013.*

[10] Howard, A. G., "Some Improvements on Deep Convolutional Neural Network Based Image Classification", *arXiv preprint arXiv:1312.5402, 2013.*

[11] Krizhevsky, A., Sutskever I., and Hinton, G., "Imagenet classification with deep convolutional neural networks", *Advances in neural information processing systems, 2012.*

[12] G. Triantafyllidis, M. Dimitriou, T. Kounalakis, and N. Vidakis, "Detection and classification of multiple objects using an RGB-D sensor and linear spatial pyramid matching," in *ELCVIA Electronic Letters on Computer Vision and Image Analysis*, vol. 12, 2013.

[13] Currency recognition system using image processing : http://www.diva-portal.org/smash/get/diva2:324361/FULLTEXT02

[14] "Artificial intelligence," in *wikipedia.org*. [Online]. Available: https://en.wikipedia.org/wiki/Artificial_intelligence#Deduction.2C_reasoning.2C_problem_solving. Accessed: Mar. 7, 2016.

[15] R. Socher, "Recursive deep learning for natural language processing and computer vision" 2014.