Scene Extraction and Recognition

Haizi Yu
Computer Science Department
Stanford University
haiziyu@stanford.edu

1. Motivation and introduction

Object detection has been a hot topic in Computer Vision for a long time. However, through tons of research papers on object detection, people tend to treat the desired object individually and separately, as a single entity. In contrast, this project will aim to detect a group of objects which tend to go together in our daily life, e.g., knife and fork, sky and cloud, in hope of exploring the relationship among objects. There comes the rough idea of scene.

Existing scene recognition techniques tend to treat an entire scene image as a training example and give it a hand coded label. The downside of this is the concept of a scene heavily relies on photographer's personal interest, thus, it is very likely that the scene is not revealed in a natural way. One direct consequence of this is that it will be hard and confusing to give an image a label, if multiple scene appear in the same image.

So, this project tends to represent each image in the training set as a union of different scene. Together with the techniques developed in object group learning, it will be possible in the end to extract any scene contained in an image and recognize it.

2. schema

2.1. Data collection

Data will be collected as a set of annotated images. From vast amount of available images, to make the learning procedure efficient, one round of pre-selection is performed, in the preference of each selected image apparently contains multiple scene (in the human intuition) instead of just one.

2.2. Object group learning

A brand new idea of learning object group will be developed and implemented. The basic idea is that we are going to pre-define a set of bases in the image concept space, i.e, rather than composed by pixels, an image is represented by object-oriented and/or scene-oriented concepts. (Note that this is how human think of a picutre.) Then each image will be conceptually represented as a linear combination of the bases. In the scope of this paper, each basis will encode the exact information of a single scene, e.g., what does the scene look like and where might it be.

The set of scene bases will be trained and learned to satisfy certain properties. Once the set is created, we can "project" the input image to each basis to get the coefficients and to reconstruct the input image in concept using the linear combination.

2.3. Scene extraction

After object group learning, we'll be able to divide a test image into different parts in terms of scene. The task here is to localize each scene in the image, retrieve it and recognize it.

3. Expected result

We are using a data driven approach to gradually build up the knowledge system in an intelligent agent's brain. So through this project, we are in the hope of Computer Vision not only has the ability to identify individual object, but also has the knowledge of understanding the relationships among objects. And this in turn, we believe, will also greatly improve identification of single object in a already understood scene.