Semantic Image Annotation Based on Bayesian Networks

Ziyu Zhou zz2637
Topic: Machine Reasoning with Large-Scale Bayesian Networks
Date: February 14. 2020

Abstract: Retrieve image from a huge database required precisely image annotation. In this project, an image annotation system will be built to use context inference. The goal is to use the low-level visual content which used by machine to represent images, to deduce the high-level concepts that used by human to describe images. The Bayesian network is used as a classifier to infer possible annotation of the image.

1. Background (Review of Related Literature):

With the using of multimedia grows, millions of image content created each, which is complex and hard to retrieve. To achieve information retrieval from those huge databases, images must be indexed and annotated precisely. Obviously, this process is hard to be done manually by human. Therefore, we want a method that can automatically annotate images. The key problem is how to use the low-level visual content which used by machine to represent images, to deduce the high-level concepts that used by human to describe images. To reduce this semantic gap, we introduce a system based on Bayesian network.

2. Introduction to the Project:

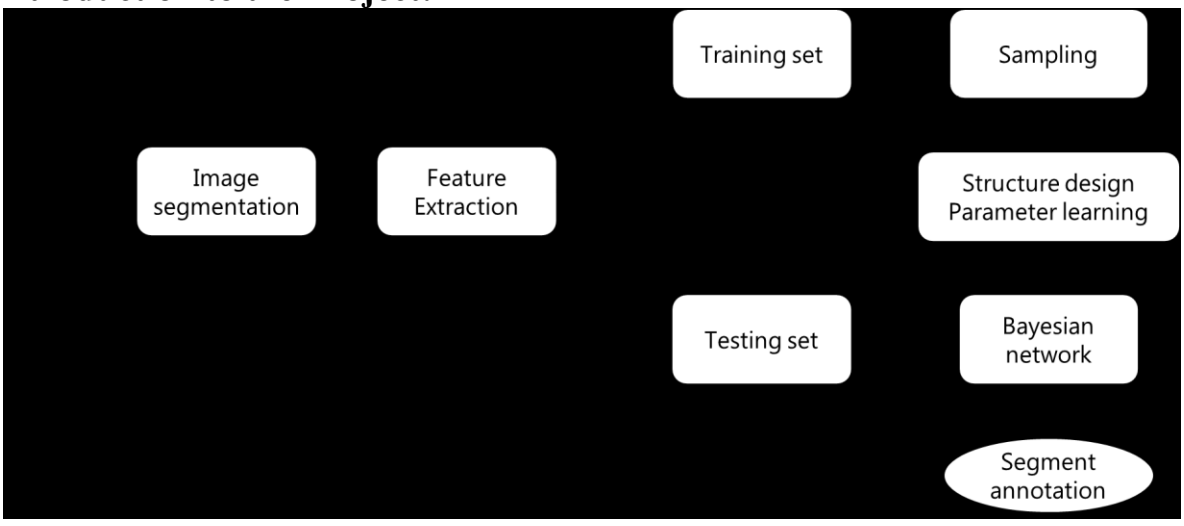

Figure 2.1 The flowchart of the project

To goal of this project is to achieve accurately image annotation. The flowchart of this project is shown below in Figure 2.1. Three main steps in this project are image segmentation, feature extraction and training the Bayesian network.
Usually, the features vector extracted from the entire image loses local information. Therefore, it is necessary to segment an image into regions or objects of interest and use of

local characteristics. In image segmentation step, some method will try some method likes K-means, threshold or region growing. The performance of each method will be evaluated. After image segmentation, the low level feature of each part of image will be extracted to training. In the feature extraction step, there are some features I plan to extract, which are texture feature, GIST, and MOO.
With the large-scale of data, it is quite impossible to train all the data. Therefore, the features will be sampled before training. In the training step, the Bayesian network will be used as a classifier. The sampling features will be used to train the network. In the testing step, the network will be used to annotate the image.

3. Introduction to the Dataset:

The dataset used in this project are ETH-80 dataset [1] and Corel-10k [2] dataset. They are two datasets with different complexity. The project will start with the simple dataset to verify the project frame and further adjust the frame by the complex one.

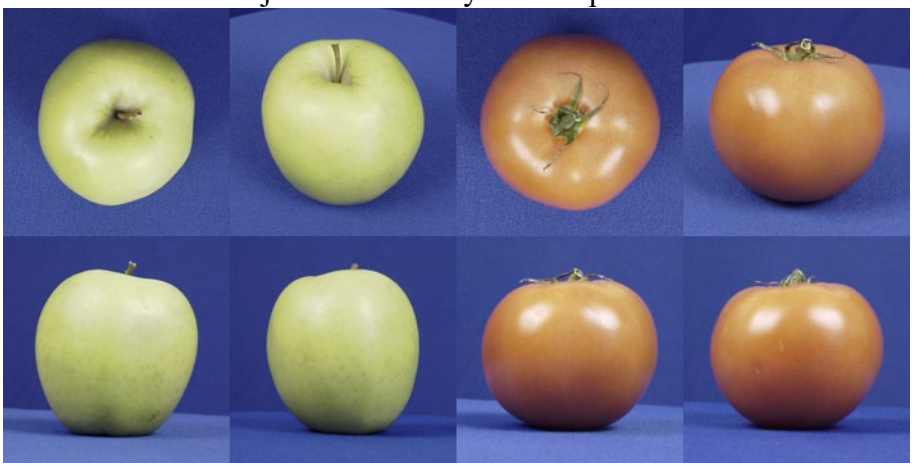

Figure 3.1 Some data from ETH-80

The ETH-80 dataset contains visual object images from 8 different categories, which are apples, cars, cows, cups, dogs, horses, pears and tomatoes. For each category, there are 10 object instances and 41 images for each object instance captured from different viewpoints. The original resolution of the images in the ETH-80 is ranging from 400×400 to 700×700 pixels, depending on object size. All images are cropped, so that they contain only the object, centered in the image, plus a 20% border area.

Figure 3.2 Some data from Corel-10k

Corel-10k dataset contains visual object images from 100 different categories, which are sunset, beach, flower, building, car, horses, mountains, fish, food, door, etc. Each category contains 100 images of size 192×128 or 128×192 pixels. All images are from real world, with complex background.

4. Plan:

Figure 4.1 Milestone planning

There are 4 stages in the project.
Before milestone 1, the main things I plan are collecting information of project, writing proposal and building the entire framework of the project.
Before milestone 2, the main things I plan are basically achieve the three main parts in the project, which are image segmentation, feature extraction and training the Bayesian network.
Before milestone 2, the main things I plan to optimize the model and try different method.
At final, the whole software will be build.

Reference:

[1] B. Leibe and, B.Schiele. Analyzing Appearance and Contour Based Methods for Object Categorization. CVPR, vol.2, pp. 409-415, 2003

[2] Guang-Hai Liu, Jing-Yu Yang, etc,. Content-based image retrieval using computational visual attention model, Pattern Recognition, 48(8) (2015) 2554-2566.