Real Estate Price Estimator

Computer Vision Final Project Report

Spring 2015

Eman Hamed

Mohamed Ghoneim

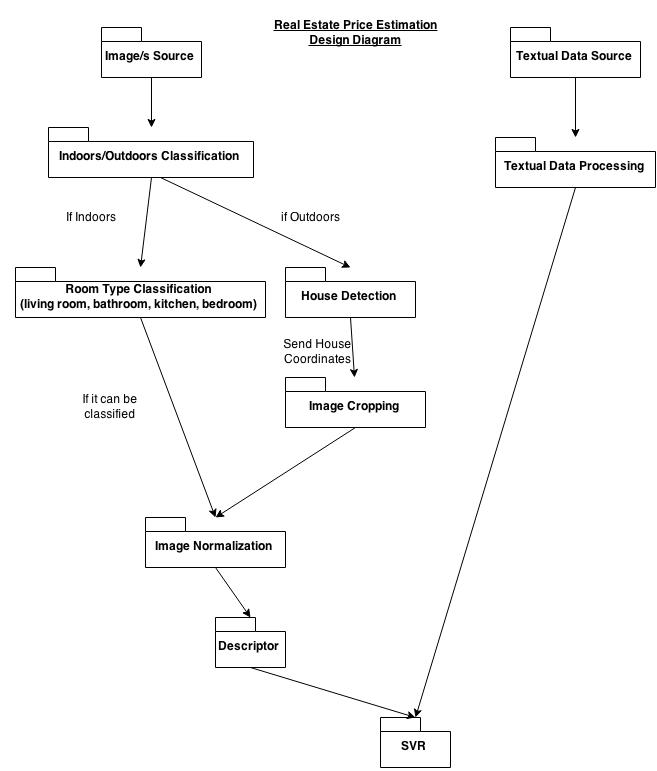
Ahmed Abdelmageid

**Project Description**

With the increasing number of sales in the real estate market, it has become very tedious work to estimates prices of properties based on information such as their geographical locations, local landmarks, and other price changing factors. Companies that works in the field of real estate sales waste hours estimating properties’ prices. This is becoming ever challenging as such companies attempt to sell properties in remote states because in that case research would have to be done regarding these areas in order to have a proper estimation to the value of each property. It would be of a great economic benefit to automate this process of price estimation to save resources.

**Project Overview**

The following figure shows the initial design of the system, which has been through some changes. This design was built under the assumption that the input information from the user might have some missing data/images. One of the inputs to the system is the frontal image of the home. Initially, it was assumed that this image might contain some noise background; however, the current design assumes that the frontal image includes the house only, without any background. Therefore,the house detection, image cropping, and image normalization modules are no longer needed as these were the modules responsible for cropping and normalizing the house from the input image.

****

The current design works as follows:

The input of the system consists of four images for each house. The images are for the bedroom, bathroom, kitchen, and a frontal outdoor image of the house. Each house is accompanied by annotations that describe the number of bedrooms, bathrooms, area, zip code, and price. These annotations are processed for all the input house images and fed to the SVR. The SVR's main task is to estimate the price of the input house. Bag of words technique is applied to the input four images from all houses in the training set, then the output histograms from the technique are forwarded to an SVM classifier which decides the type of the image (whether it is bathroom, bedroom, kitchen, or the frontal image). These histograms are also fed to the main SVR which along with the houses annotations would estimate the price of the input house.

**Results and Screenshots**

**Challenges**

There are several challenges that were faced implementing this project. One of the main challenges is the lack of a proper dataset that is big enough to train the Bag of Words algorithm, and the SVR. Therefore, the dataset had to be collected manually. About 70 houses were collected and arranged in a dataset where each house had four images (one for each of the bedrooms, bathrooms, kitchen, and the frontal image of the house), long with textual data that describe the number of bedrooms and bathrooms in each house, as well as, its area, zip code, and price. The relatively small number of houses in the dataset has lead to insufficient information being given to the SVR. As a result, the overall accuracy of the SVR is below average.

Initially, there was an extra step done to the input images before being fed to the Bag of Words algorithm. The input images were classified into indoors (bedrooms, bathrooms, and kitchen images) and outdoors (the frontal image). However, the accuracy of this Naive Bayesian Classifier was about 80%. Consequently, the bag of words algorithm had some images of outdoor images while it is supposed to only have images of bedrooms, bathrooms, and kitchens. This has lead to poor results from the Bag of Words algorithm, and this problem has forwarded to the main SVR. In order to solve this problem, the indoor/outdoor classification step was ignored and all four images were given to the Bag of Words algorithm. This assumes the frontal image of the house to be an extra room type.