



Faculty of Computers & Information



Cairo University

Master Thesis

Content-based Image Retrieval for Agricultural Crops

Submitted by

Esraa Mohammed Hashem Elhariri

Submitted to

Faculty of Computers and Information, Cairo University

in partial fulfillment of the requirements for M.Sc. degree in computer
science

Faculty of Computers and Information, Cairo University

Giza, Egypt

February 2015



Faculty of Computers & Information



Cairo University

Master Thesis
Content-based Image Retrieval for Agricultural Crops

Submitted by

Esraa Mohammed Hashem Elhariri

Submitted to

Faculty of Computers and Information, Cairo University

in partial fulfillment of the requirements for M.Sc. degree in computer
science

Supervised By

Prof.Dr. Amr Badr

Professor at Computer Science Department -
Cairo University

Prof.Dr. Aboul Ella Hassanien

Professor at Information Technology
Department - Cairo University

Prof.Dr. Nashwa ElBendary

Associate professor at Arab Academy for Science, Technology, and Maritime Transport



Faculty of Computers & Information



Cairo University

Abstract

Fruits and vegetables quality is one of the most important factors that helps ensuring a consistent marketing of fruits and vegetables. As ripeness is the main indicator for fruits and vegetables quality from customers perspective, the determination of fruits and vegetables ripeness stages is a basic industrial concern regarding crops production in order to get high quality product. Automatic ripeness evaluation of fruits and vegetables is an essential research topic as it may prove benefits in ensuring optimum yield of high quality product, this will increase the income because agriculture crops play a crucial role in the life of an economy. It is the backbone of the country's economic system as it is one of the prime sources of the countries' national income. Identifying physiological maturity and harvest maturity of a agricultural crops correctly will ensure timely harvest to avoid cutting of either under- and over-ripe agricultural crops, so monitoring and controlling agricultural crops ripeness is a very important issue in the agricultural crops industry since the state of ripeness during harvest, storage and market distribution determines the quality of the final product measured in terms of customer satisfaction.

Doing this manually will be time consuming, cost high ,needs a lot of work and efforts and may be inaccurate because of inconsistent and subjective of manual human grading technique. This thesis presents a content-based image classification approach to automate the ripeness assessment process of tomato/ bell pepper via examining and classifying the different ripeness stages. Datasets of total 250 and 175 images for tomato and bell pepper respectively were used for both training and testing datasets with 10-fold cross-validation. Training dataset is divided into 5 classes representing the different stages of tomato and bell pepper ripeness.

The proposed system consists of three phases; namely pre-processing, feature extraction, and classification phases. During pre-processing phase, the proposed system resizes images to 250x250 pixels, in order to reduce their color index, and



Faculty of Computers & Information



Cairo University

the background of each image will be removed using background subtraction technique. Also, each image is converted from RGB to HSV color space. For feature extraction phase, Principal Component Analysis (PCA) algorithm is applied in order to generate a feature vector for each image in the dataset. The proposed system depends totally on color features (colored histogram and color moments), since the surface color of tomato/ bell pepper is the most important characteristic to observe ripeness. Finally, for classification phase, the proposed system applied different classification algorithms for classification of different ripeness stages.

This thesis presents many different scenarios for ripeness classification of tomato and bell pepper such as, One-against-All SVM multi-class, One-against-One SVM multi-class, Random Forests (RFs) and Linear Discriminant Analysis (LDA) classification algorithms. Also this thesis presents a comparative analysis between different algorithms.