Computers and Electronics in Agriculture 70 (2010) 96–104

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Automatic fruit and vegetable classification from images

Anderson Rocha[a,*], Daniel C. Hauagge[b], Jacques Wainer[a], Siome Goldenstein[a]

[a] Institute of Computing, University of Campinas (Unicamp), Campinas, Brazil
[b] Department of Computer Science, Cornell University, Ithaca, United States

ARTICLE INFO

Article history:
Received 17 April 2008
Received in revised form 28 August 2009
Accepted 14 September 2009

Keywords:
Feature and classifier fusion
Multi-class from binary
Automatic produce classification
Image classification

ABSTRACT

Contemporary Vision and Pattern Recognition problems such as face recognition, fingerprinting identification, image categorization, and DNA sequencing often have an arbitrarily large number of classes and properties to consider. To deal with such complex problems using just one feature descriptor is a difficult task and feature fusion may become mandatory. Although normal feature fusion is quite effective for some problems, it can yield unexpected classification results when the different features are not properly normalized and preprocessed. Besides it has the drawback of increasing the dimensionality which might require more training data. To cope with these problems, this paper introduces a unified approach that can combine many *features* and *classifiers* that requires less training and is more adequate to some problems than a naïve method, where all features are simply concatenated and fed independently to each classification algorithm. Besides that, the presented technique is amenable to continuous learning, both when refining a learned model and also when adding new classes to be discriminated. The introduced fusion approach is validated using a multi-class fruit-and-vegetable categorization task in a semi-controlled environment, such as a distribution center or the supermarket cashier. The results show that the solution is able to reduce the classification error in up to 15 percentage points with respect to the baseline.

1. Introduction

Recognizing different kinds of vegetables and fruits is a recurrent task in supermarkets, where the cashier must be able to point out not only the species of a particular fruit (i.e., banana, apple, pear) but also its variety (i.e., Golden Delicious, Jonagold, Fuji), which will determine it's price. The use of barcodes has mostly ended this problem for packaged products but given that consumers want to pick their produce, they cannot be packaged, and thus must be weighted. A common solution to this problem is issuing codes for each kind of fruit/vegetable; which has problems given that the memorization is hard, leading to errors in pricing.

As an aid to the cashier, many supermarkets issue a small book with pictures and codes; the problem with this solution is that flipping over the booklet is time-consuming.

This paper reviews several image descriptors in the literature and introduces a system to solve the problem by adapting a camera to the supermarket scale that identifies fruits and vegetables based on color, texture, and appearance cues.

Formally, given an image of fruits or vegetables of only one variety, in arbitrary position and number, the system must return a list of possible candidates of the form (species, variety). Sometimes, the object can be inside a plastic bag that can add specular reflections and hue shifts.

Given the variety and the impossibility of predicting which kinds of fruit/vegetables are sold, training must be done on-site by someone with little or no technical knowledge. Therefore, the system must be able to achieve a high level of precision with only a few training examples (e.g., up to 30 images).

Often, one needs to deal with complex classification problems. In such scenarios, using just one feature descriptor to capture the classes' separability might not be enough and feature fusion may become necessary.

Although normal feature fusion is quite effective for some problems, it can yield unexpected classification results when the different features are not properly normalized and preprocessed. Besides it has the drawback of increasing the dimensionality of the data which might require more training examples.

This paper presents a unified approach that can combine many *features* and *classifiers*. It requires less training and is more adequate to some problems than a naïve method, where all features are simply concatenated and fed independently to each classification algorithm. We expect that this solution will endure beyond the problem solved in this paper.

The introduced fusion approach is validated using an image data set collected from the local fruits and vegetables distribution center and made public. The image data set contains 15 produce categories

* Corresponding author at: Av. Albert Einstein, 1251, Cx. Postal 6176, CEP 13083-970, Campinas/SP, Brazil. Tel.: +55 19 8801 0992; fax: +55 19 3521 5888.
E-mail address: anderson.rocha@ic.unicamp.br (A. Rocha).
URL: http://www.ic.unicamp.br/~rocha (A. Rocha).

0168-1699/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.compag.2009.09.002

comprising 2633 images collected on-site in a period of 5 months under diverse conditions. The implemented solution achieves a classification error less than 2% for the top one responses. With the top two responses such error is smaller than 1%.

Section 2 gives a brief overview of previous work in object recognition and image categorization. Section 3 presents the different kinds of image descriptors used in this paper as well as the produce data set. Section 4 introduces the solution for feature and classifier fusion, and Section 5 presents experimental results. Finally, Section 6 draws the conclusions and future directions.

2. Literature review

Recently, there has been a lot of activity in the area of *Image Categorization*. Previous approaches considered patterns in color, edge and texture properties (Stehling et al., 2002; Unser, 1986; Pass et al., 1997); low- and middle-level features to distinguish broad classes of images (Rocha and Goldenstein, 2007; Lyu and Farid, 2005; Cutzu et al., 2005; Serrano et al., 2004). In addition, Heidemann (2004) has presented an approach to establish image categories automatically using histograms, colors and shape descriptors with an unsupervised learning method.

With respect to the produce classification problem, *Veggievision* (Bolle et al., 1996) was the first attempt of a Supermarket Produce recognition system. The system uses color, texture and density (thus requiring extra information from the scale). However, as this system was created sometime ago, it does not take advantage of recent developments. The reported accuracy was $\approx 95\%$ in some scenarios but to achieve such result it uses the top four responses. The data set used in this paper is more demanding in some respects; while the data set in Veggievision had more classes, the image capturing hardware gave a more uniform color and suppressed specular lights. The data set assembled in this paper has greater illumination and color variation among images, also there is no measure to suppress specularities.

In general, the produce classification problem can be seen as a special instance of object's categorization. Turk and Pentland (1991) employed principal component analysis and measured the reconstruction error of projecting the image to a subspace and returning to the original image space. We believe this is ill suited for produce classification because it depends heavily on illumination, pose and shape.

Recently, Agarwal et al. (2004) and Jurie and Triggs (2005) adopted approaches that break down the categorization problem to the recognition of specific parts that are characteristic of each object class. These techniques, generally called bag-of-features (Marszalek and Schmid, 2006; Grauman and Darrel, 2005; Sivic et al., 2005), showed promising results even though they do not try to model spatial constraints among features.

Weber (2000) takes into account spatial constraints using a generative constellation model. The algorithm can cope with occlusion in a very elegant manner, albeit very costly (exponential in the number of parts). A further development made by Fei-Fei et al. (2006) introduced prior knowledge into the estimation of the distribution, thus reducing the number of training examples to around 10 images while preserving a good recognition rate. Even with this improvement, the problem of exponential growth with the number of parts persists, which makes it unpractical for the problem presented in this paper, which requires speed for on-line operation.

Another interesting technique was proposed by Berg et al. (2005). In that work, feature points are found in a gradient image. The points are connected by a joining path and a match is signalized if the found contour is similar enough to the one in the database. A serious drawback of this method for produce classification is that it requires a nonlinear optimization step to find the best contour; besides that it relies too heavily on the silhouette cues, which are not a very informative feature for fruits like oranges, lemons and melons.

3. Materials and methods

In general, image categorization relies on combinations of statistical, structural and spectral approaches. Statistical approaches describe the objects using global and local descriptors such as mean, variance, and entropy. Structural approaches represent the object's appearance using well-known primitives such as patches of important parts of the object. Finally, spectral approaches describe the objects using some spectral space representation such as Fourier spectrum (Gonzalez and Woods, 2007).

This paper analyzes statistical color and texture descriptors as well as structural appearance descriptors to categorize fruits and vegetables in a multi-class scenario. Since the best combination of features was not known for this problem, we analyze several state-of-the-art Computer Vision features in many different ways, and assemble a system with good overall accuracy using underpinned cross-validation procedures that allows the combination of the best features and classifiers into a single and unified approach.

The following sections present the statistical and structural descriptors used in this paper, as well as the data set assembled for the validation process.

3.1. Supermarket Produce data set

The *Supermarket Produce* data set is one of the contributions in this paper.[1] In general, there are a few well-documented image data sets available for image categorization and content-based image retrieval tasks for testing algorithm performance. ALOI[2] and Caltech[3] are two examples of such data sets for general categorization.

The *Supermarket Produce* data set is the result of 5 months of on-site collecting in the local fruits and vegetables distribution center.

The images were captured on a clear background at the resolution of 1024×768 pixels, using a Canon PowerShot P1 camera. For the experiments in this paper, they were downsampled to 640×480. The data set comprises 15 different categories: Plum (264), Agata Potato (201), Asterix Potato (182), Cashew (210), Onion (75), Orange (103), Taiti Lime (106), Kiwi (171), Fuji Apple (212), Granny-Smith Apple (155), Watermelon (192), Honeydew Melon (145), Nectarine (247), Williams Pear (159), and Diamond Peach (211); totalizing 2633 images. Fig. 1 depicts some of the classes of the data set.

All of the images were stored in RGB color-space at 8 bits per channel. The images were gathered at various times of the day and in different days for the same category. These features increase the data set variability and represent a more realistic scenario. Fig. 2 shows an example of Kiwi and Granny-Smith Apple categories with different lighting. The differences are due to illumination, no image pre-processing was performed.

The *Supermarket Produce* data set also comprises differences in pose and in the number of elements within an image. Fig. 3 shows examples of the Cashew category. Note that there are variations in the pose of the Cashew's plastic repository. In addition, Fig. 4 shows the variability in the number of elements within an image.

Sometimes, the elements are inside a plastic bag which adds specular reflections to the analyzed image. Furthermore, the pres-

[1] Freely available from http://www.liv.ic.unicamp.br/~undersun/pub/communications.html.

[2] http://staff.science.uva.nl/~aloi.

[3] http://www.vision.caltech.edu/Image_Datasets/.