Some Fuzzy Tools for Evaluation of Computer Vision Algorithms

Andrey Osipov, Scientific-Research Institute of System Analysis, Russian Academy of Sciences (SRISA RAS), Moscow, Russia

ABSTRACT

In this article, some issues related to the performance evaluation of computer vision algorithms within
the version of direct empirical supervised evaluation method developed at SRISA RAS are considered.
This approach partly relies on the elements defined by using the fuzzy set theory, in particular, fuzzy
similarity measures and fuzzy reference ground truth images. Some known measures of segmentation
quality are considered and their extensions, representing the fuzzy similarity measures, are offered.
As an example, the author considers an application of fuzzy ground truth images and fuzzy similarity
measures, including some newly introduced ones, to the evaluation of face recognition algorithms.

KEYWORDS

Face Recognition, Fuzzy Similarity Measures, Ground Truth Images, Image Segmentation, Performance Evaluation

INTRODUCTION

For more than half a century of computer vision research thousands of various algorithms were
proposed. Many of them have multiple software implementations (e. g. the famous Canny edge
detector). As a result, the developer of computer vision system faces a complicated task of choosing
the most appropriate algorithms for his specific purposes. For many reasons (see e. g. Wirth et. al.,
(2006)), evaluation of image processing and analysis algorithms for practical purposes has no unique
method. To date, several attempts to classify the existing evaluation methods have been made. In
particular, the following classification of evaluation methods for image segmentation has been offered
in Zhang et. al., (2008):

1. Subjective evaluation;
2. Objective evaluation:
   a. System level evaluation;
   b. Direct evaluation:
      i. Analytical methods;
      ii. Empirical methods.

DOI: 10.4018/IJCVIP.2018010101

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
In turn, the empirical methods can be divided into supervised and unsupervised ones. In principle, such classification is appropriate to classify evaluation methods for another class of computer vision algorithms (e.g. edge detectors).

Supervised methods are also known as empirical discrepancy methods. The latter definition is probably more appropriate, since such methods perform a comparison between a processed image (algorithm’s output) against a reference image which is often referred to as a ground-truth, by using some quantitative evaluation criteria - similarity measures. The ground truth images are often manually created and contain the features which are ideal from the evaluator’s viewpoint. For instance, if we evaluate the edge detectors, then for every test image there is a matching ground truth image containing ideal (user-defined) edges.

For the purpose of empirical supervised evaluation of image processing and analysis algorithms, we have developed the software system named PICASSO. Originally it was designed to compare edge detection algorithms on a set of artificial 2D images. Its current version evaluates a wider range of algorithms including e.g. image segmentation algorithms, image restoration methods, texture analysis algorithms, (see Gribkov et. al., (2005)). Also, the testing technique has been improved. In particular, now it includes some elements of fuzzy logic. This inclusion is justified by the growing amount of image processing and analysis methods which rely on the fuzzy set theory (see Bezdek et. al. (1999) and Tizhoosh & Haussbecker, (2000)). For example, many PCA-based face recognition methods (see Yang et. al., (2010) and references therein) are using the fuzzy k – nearest neighbor algorithm offered in Keller et. al., (1985) to build up scatter matrices. Also, in processing of remote sensing images, due to insufficient resolution of the sensor, often it is difficult to assign some pixels to one pure class (e.g. to the “forest”, “water”, or “urban land”). This uncertainty has led to the idea of using elements of fuzzy set theory for handling such tasks (a thorough discussion is contained in Lu & Weng, (2007)). Obviously, the comparative evaluation technique must be able to handle these “fuzzy” methods and to compare simultaneously both “fuzzy” and “non-fuzzy” algorithms.

Note that methods of comparative evaluation of these two types of algorithms have not been developed until now. In some papers about the analysis of remote sensing images, several statistical performance measures were generalized for fuzzy case. In particular, these measures allow one to compare fuzzy sets with respect to their crisp counterparts. Such measures were called by Jäger and Benz (Jäger & Benz, (2000)) as fuzzy similarity measures. Also, in their paper the concept of a fuzzy ground truth image was discussed. Although the paper was directed at aerial image classification, some of its results can be applied as well to comparative evaluation of edge detectors. This observation resulted in a paper by Osipov (2007). In this paper, the fuzzy ground truth images used for testing edge detection algorithms were identified with membership functions of edge class, which for each pixel of a given test image takes the values from 0 to 1. The ordinary ground truth images containing the reference edge maps were identified with the characteristic functions of the pixels forming these edge maps (taking the value 1 for edge pixels and 0 otherwise).

One of the features of our evaluation method is that depending on the specific property of the evaluated algorithm, different fuzzy ground truths corresponding to the same test image can be used. For example, one fuzzy reference image can be used to test the detection of weak edges, and another one - to test the ability to generate continuous edge contours (its membership function is sensitive to the gaps in the edge map). Also, our experiments have shown that these fuzzy ground truths can be used to study the ability of edge detectors to find image feature points (for example, the corner points of a rectangular). Assigning the higher values of the membership functions to these pixels compared to the other pixels on the ground truth edge map, we get the higher values of fuzzy similarity measures when such pixels were marked as edges by the evaluated algorithm. For the quantitative evaluation, we used in this paper the fuzzy similarity measures introduced in Jäger & Benz, (2000). However, it was noted that these measures cannot provide a comprehensive evaluation of the algorithms’ performance, e.g. been the statistical measures by their nature, they are unable to provide a proper evaluation of the localization performance of the edge detectors. Note, that as mentioned by Canny, there is a sort
Software Framework of Medical Visualization Algorithms

www.igi-global.com/chapter/software-framework-medical-visualization-algorithms/30630?camid=4v1a