J. Vis. Commun. Image R. 32 (2015) 156–169

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

A comparison of interest point and region detectors on structured, range and texture images ☆

Wajahat Kazmi *, Hans Jørgen Andersen

Department of Architecture, Design and Media Technology, Aalborg University, Rendsburggade 14-5, 9000 Aalborg, Denmark

ARTICLE INFO

Article history:
Received 30 July 2014
Accepted 4 August 2015
Available online 10 August 2015

Keywords:
Affine invariant regions
Edge shapes
Local features
Detectors
Descriptors
Image retrieval
Harris corners
Hessian
SIFT
SURF

ABSTRACT

This article presents an evaluation of the image retrieval and classification potential of local features. Several affine invariant region and scale invariant interest point detectors in combination with well known descriptors were evaluated. Tests on building, range and texture databases were carried out in order to understand the effects of the nature and the variability of the data on the performance of the detectors in terms of their invariance to affine deformations and scale changes. Furthermore, a novel multi-scale edge shape detector, Twin Leaf Regions (TLR) is also proposed using a graph based image decomposition. In TLR, Affine adaptation is avoided in order to reduce the offset from the edges so that pure edges shapes are captured in multiple scales. In the evaluation of building recognition, both homogeneous affine regions (such as Maximally Stable Extremal Regions (MSER)) and corner based detectors (such as Hessian and Harris with both Affine/Laplace variants, SURF with determinant of Hessian based corners and SIFT with difference of Gaussians) acquired more than 90% mean average precision, whereas on range images, homogeneous region detector did not work well. TLR offered good performance than MSER and comparable performance to Harris Affine and Harris Laplace in range image classification and texture retrieval. But its performance was low in building recognition. In general, it was observed that the affine and scale invariance becomes less effective in range and textured images. It is also shown that in a bi-channel approach, combining surface and edge regions (MSER and TLR) boosts the overall performance. Among the descriptors, SIFT and SURF generally offer higher performance but low dimensional descriptors such as Steerable Filters follow closely.

1. Introduction

Local features in computer vision outperform global features in the presence of large view point changes, clutter and occlusion [1]. Several local detectors and descriptors have been introduced in the literature and the problems such as object recognition have been addressed to a fairly good extent. However, the performance is subject to the application at hand. The algorithms, by large, address general computer vision data which are 2D color or grayscale images of scenes with either structured (man made objects constituted from simpler geometrical shapes) or a combination of structured and natural objects usually undergoing rigid transformations.

Images containing slightly different forms of data such as 3D or depth images in which color or texture is not present and the object's shape is what defines all the shades may pose a slightly different challenge [2–4]. Another case is of the objects with smooth boundaries where the conventional interest point detection does not work very well [5]. In a textured scene, the concept of an interest point, a rigid object or transformation may be less appealing. Even more challenging are the images of deformable objects such as the outdoor images of plants, for example, with several leaves under the effects of biological growth, wind and light. In such cases, distinctive shape features can be obtained from edges and the shapes of their contours.

1.1. Features from edge fragments

In a typical feature extraction process, interest points that lie along an edge are usually ignored because they offer low repeatability [8]. On the other hand, edges are high gradient neighborhoods and carry important information regarding the shape of the objects [9]. In the absence of colors and texture, edges become increasingly important as is the case with 3D or depth data. Owing to this reason, edge contours have been successfully used in object

☆ This paper has been recommended for acceptance by Prof. M.T. Sun.

* Corresponding author.
E-mail addresses: wajahat.kazmi@outlook.com (W. Kazmi), hja@create.aau.dk (H.J. Andersen).

http://dx.doi.org/10.1016/j.jvcir.2015.08.004
1047-3203/

classification. Mikolajczyk *et al.* [9] used edge points with a Laplacian based measure for edge support selection to incorporate scale invariance. Shotton *et al.* [10] randomly selected few edge fragments and pruned them by intensity based clustering forming a code book of contours. Ferrari *et al.* [11] used chains of connected contour segments for the object class detection. A more common approach is using sampled edge points [12]. The concept of BoB (Bag-of-Boundaries) introduced by Arandjelovic and Zisserman [5] matched the post-segmentation boundaries for retrieval of objects with low surface texture and smooth boundaries. For a machine vision system, the worst deformations are projective transformations and developing robustness to them cannot be error free. Edge sampling, unfortunately, do not provide invariance to such deformations. The best compromise can be reached with affine invariant regions.

1.2. Affine invariant region detectors

The *Affine Invariant Regions* provide partial invariance to projective transformations [13,14]. Affine regions target different shape features of an object, such as Maximally Stable Extremal Regions (MSER) [15] seek the bounds of homogeneous stable regions following a watershed approach while Edge Based Regions (EBR) [14] find seed points from Harris corners and follow the edges along the two directions of the corner to trace the boundary of the edges.

The performance of affine regions has been remarkably good [16]. Many recent algorithms perform slightly better, such as Wαsh detector [17], Medial Feature Detector (MFD) [18] and Boundary Preserving dense Local Regions (BPLR) [19]. Mikolajczyk and Schmid [13] proposed Harris Affine and Hessian Affine regions which originate from Harris and Hessian corners, respectively, and occupy the affine covariant neighborhood in a multi-scale manner. Therefore, they tend to partially capture edges as shown in Fig. 1(g) and (i), which seems to be another way to include edge fragments with robustness against affine deformations. But they do not take into account the edge contours which limits the information content related to its shape (Fig. 1 shows some of the commonly used affine and scale invariant detector on a plant image).

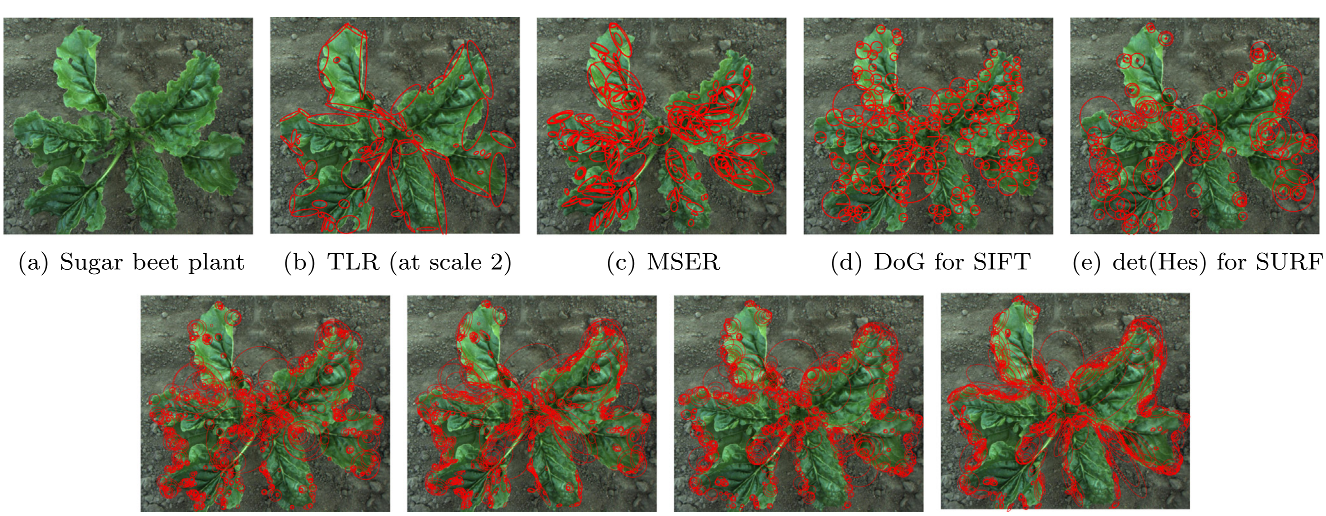
The reason is that, in the process of finding a neighborhood covariant to affine transformations, the region (*distinguished region* [15,16]) is iteratively mapped onto an ellipse (*measurement region*). In this process, the shape of the boundary contributing to the distinguished region could be lost, at least partially (Fig. 1(g) and (i)). Therefore, as argued by [9], the invariance to affine geometric or photometric deformations comes at the cost of the information content of the local features. If a fragment of the edge or boundary which holds a clue to its geometry is enclosed in the ellipse representing the measurement region, the extracted descriptor may also include features for the edge shape. But this is more a matter of chance than the design.

The non-affine counterparts, the Harris Laplace and Hessian Laplace (Fig. 1(f) and (h)), although avoid affine adaptation, but are limited to corners. They both start from corners and search for an extrema of Laplacian which signifies optimal scales, indirectly including some part of the edges and hence the subsequent descriptor partly addresses edge shapes. Mikolajczyk *et al.* [16] in their comparison of affine regions, point towards the need for developing affine regions for object boundaries.

Objectives

The local feature evaluations in literature usually focus on either certain type of detectors [16], descriptors [20], or data such as 3D [4]. In this article, our objective is to assess the complexity of the features w.r.t that of the data. We evaluate multi-scale, scale invariant and affine invariant detectors on three publicly available datasets including building structures, depth images and textures with a number of descriptors both for shape and texture. These data and feature types are chosen so that a comprehensive comparison can be established to facilitate an application oriented reader in understanding the pros and cons of making a choice.

The end goal of our research is to prepare tools for plant leaf recognition for which the edge shapes are an important discriminant. Since affine or scale invariant detectors do not exclusively focus on edge shapes, therefore, we also elaborate the details of a graph based method for multi-scale interest region detection introduced in [21] followed by its thorough evaluation. The idea is to avoid the affine adaptation and focus on pure edge shapes similar to edge sampling, but in a multi-scale manner. The underlaying hypothesis is that it would be a better alternative to edge sampling as a multi-scale support approximates affine transformations locally [9].



(a) Sugar beet plant (b) TLR (at scale 2) (c) MSER (d) DoG for SIFT (e) det(Hes) for SURF

(f) Harris Laplace (g) Harris Affine (h) Hessian Laplace (i) Hessian Affine

Fig. 1. Sample sugar beet plant with selected local features (*Note:* For better viewing, the number of regions shown is lesser than actually detected. Background was removed before applying the detectors by using Excess Green Index [6] and Otsu thresholding [7]).