Neurocomputing 173 (2016) 1422–1435

Contents lists available at ScienceDirect

Neurocomputing

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

Scale invariant representation of imbalanced points

Qi Li*, Yongyi Gong

Cisco School of Informatics, Guangdong University of Foreign Studies, China

ARTICLE INFO

Article history:
Received 22 February 2014
Received in revised form
25 August 2015
Accepted 5 September 2015
Communicated by Hantao Liu
Available online 25 September 2015

Keywords:
Scale
Interest point
Imbalanced point
Matching

ABSTRACT

Imbalance oriented candidate selection was introduced as an alternative of non-maximum suppression, aiming to improve the localization accuracy. To distinguish interest points detected via non-maximum suppression, we call interest points detected via imbalance oriented selection *imbalanced points*. Scale assignment for imbalanced points is not straightforward because of a dilemma of involving non-maximum suppression. The scale space theory, a popular scale assignment scheme, requests non-maximum suppression to detect extreme points from scale spaces, while imbalanced points are expected to be free of non-maximum suppression in order to maintain the localization accuracy. In this paper, we propose a bypass strategy that circumvents the above dilemma by establishing an association between an imbalanced point and a certain interest point with a known scale (e.g., Lowe's keypoints and Hessian–Laplace). Furthermore, we propose a hybrid representation of imbalanced points for a two-layer matching scheme, where the first-layer matching is based on discriminant SIFT-type descriptors of imbalanced points, and the second-layer matching is based on patch-type descriptors. Experiments show the effectiveness of the proposed scale assignment and hybrid representation.

1. Introduction

Non-maximum suppression has been a popular candidate selection approach for interest point (keypoint) detection [12,31,27,28,3,30,37,6]. Non-maximum suppression may introduce localization uncertainty if an extreme point is sensitive to an imaging condition, such as camera orientation and illumination. Although the localization uncertainty appears to be small (less than $2\sqrt{2}$ pixels if a 3×3 suppression window is applied), it can have significant impacts on the accuracy of certain higher-level computations, such as the estimation of epipolar geometry, an important problem in 3D vision [4,14,33].

To improve the localization accuracy, imbalance oriented candidate selection [22] was introduced as an alternative of non-maximum suppression. (Details of imbalance oriented selection will be reviewed in Section 2). An important property of imbalance oriented selection is the *locality*, i.e., interest points can be contiguous to others in terms of 8-connectivity. For convenience, we call interested points detected via imbalance oriented selection *imbalanced points*. However, one important issue that has not been addressed in previous studies of imbalanced point detection is the scale invariant representation of these points, although experiments have shown that imbalanced points have high localization repeatability across scale variations [19]. It is worth noting that there are some other alternatives of non-maximum suppression, besides imbalance oriented selection. For example, Brown et al. [7] proposed adaptive non-maximum suppression that aims to enforce a more uniform distribution of interest points in an image plane. Adaptive non-maximum suppression has been shown to be competitive to the standard one in image mosaicing. Tuytelaars [35] proposed dense interest points, where one starts from densely sampled patches yet optimize their position and scale parameters locally. Experiments showed that dense interest points cannot reach the same localization accuracy as standard interest points.

Scale space theory [24,25] has been successfully integrated into interest point detection [28,27,11,9,1,26] to detect *scale-invariant interest points* with scales. The basic idea of scale-invariant interest point detection is to detect scale-space extreme points by non-maximum suppression, which contains region information. From this perspective, scale-invariant interest points can be more precisely called *interest regions* (invariant to scale variations) [29]. Dorko and Schmid [8] proposed a scale selection method aiming to maximize the stability of description. Instead of using the function of interest strength in the scale space, the authors derived a function of the Euclidean distance of SIFT descriptors in two consecutive scales, and consider the scale associated with the minimum of the function as the characteristic scale. It is also interesting that a recent work proposed a Fourier Transform based

* Corresponding author.
E-mail address: zsuliqi@gmail.com (Q. Li).

http://dx.doi.org/10.1016/j.neucom.2015.09.014
0925-2312/

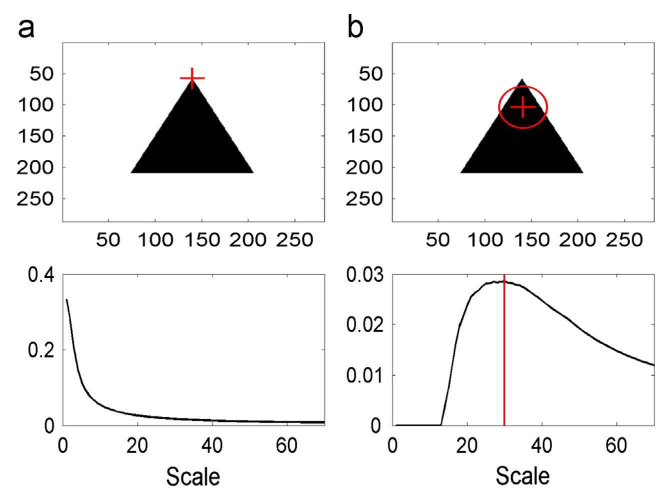

Fig. 1. Scale-response functions of a corner and a non-corner point, where a scale is in the pixel unit. The left function does not have any extrema; the right function has one extrema that determines a characteristic scale for the non-corner point. A red cross (+) denotes a corner/non-corner point, and the red circle denotes a region with the characteristic scale.

approach to construct scale invariant features without scale selection [18]. But this work did not address the localization accuracy, and it thus might not fit into the context of interest point detection. Hassner et al. [15] proposed the SLS (Scale-Less SIFT) subspace based on a subspace-to-point projection [2] in the context of dense correspondence of two images. Similar to PCA subspace, SLS subspace is learned via an unsupervised strategy.

The challenge of detecting imbalanced points with scales happens to come from its characteristics— the imbalance oriented candidate selection. Specifically, imbalance oriented candidate selection was originally introduced as an alternative of non-maximum suppression to reduce the uncertainty of interest points detected via the latter scheme. However, non-maximum suppression will be necessary to detect scale-space extreme points for the detection of imbalanced points with scales, which conflicts the rationale of imbalance oriented selection. Another intuition on the challenge of detecting imbalanced points with scales comes from their corner nature.

Fig. 1 illustrates a key difference between a corner point and a non-corner point in terms of their scale-response functions, where the scale ranges from 1 to 70 (in pixel unit). For the corner point, its scale-response function does not have any extrema inside the scale range. For the non-corner point, its function has one extrema that determines a characteristic scale (about 30).

To solve the above dilemma of involving the non-maximum suppression during the scale selection for imbalanced points, we propose a bypass strategy[1] that aims to establish a stable association between an imbalanced point and a certain interest point with a known scale (contributed by existing scale-invariant point detector/regions, e.g., Lowe's keypoints or Hessian–Laplace regions), and assigns the scale of the associated point as the scale of an imbalanced point. The rationale of the bypass strategy for the scale selection for imbalanced points is the hypothesis that corner points and regions have intrinsic relationship. So if an existing scale-invariant point/region detector can provide high localization and surface repeatability [28] of detected points/regions, and imbalanced points with bypass scales are expected to have high surface repeatability, based on their high localization repeatability.

[1] A short version was published in a conference [20].

Based on the locality property, a two-layer scheme was proposed for matching imbalanced points [23], where the first layer (also called the global layer) establishes correspondence between localities, and then the second layer (also called the local layer) refines the locality correspondence to point correspondence. It is intuitive that locality correspondence, with more local information, is more robust with respect to mismatching than conventional point correspondence. In the context of stereo correspondence, the two-layer matching scheme performs a "divide-and-conquer" strategy to address mismatching and imprecise matching separately. (Note that mismatching and imprecise matching are the two main challenges in stereo correspondence [38,14].) An important problem in the two-layer matching scheme is on how to measure the similarity between localities that is equivalent to the problem of measuring similarity of two sets of vectors [23]. Several methods have been proposed to measure similarity of localities, where a similarity measure may or may not be symmetric [19,23,21].

Given an imbalanced point, its local patch, i.e., a window of intensities of its neighboring points, was used as the representation (descriptor) of the point for the two-layer matching scheme [19,23,21]. The patch-type representation was observed to perform more effectively in the application of the estimation of the fundamental matrix of two stereo images whose baseline is relatively small [21], being compared with SIFT-type representation [27]. But it is also well known that the performance of patch-type representations is poor if images contain significant scaling or affine variations.

It is known that patch-type representations of interest points are generally more sensitive to their localization accuracy than SIFT-type representation. In other words, the patch-type representation of an interest point can be significantly dissimilar to the patch-type representation of a neighboring interest point, while the SIFT-type representation is changed less significantly due to their statistical construction. As specified by Lowe in [27], SIFT can tolerate up to 4-pixel shift due to the design of 4×4 window. This difference leads to the following consequences and tradeoff between the two presentations:

- The patch-type representation tends to perform worse than the SIFT-type representation when imaging variations are significant.
- The patch-type representation tends to contain fewer instances of imprecise matching due to the low tolerance of inaccurate localization.
- Feature extraction methods, such as Principal Component Analysis (PCA) [16,17] and Linear Discriminant Analysis (LDA) [5,34], are expected to be more effective to the SIFT-type presentation than the patch-type representation. Note that these methods assume Gaussian distribution of feature vectors, and the statistical characteristics of SIFT-type representation is more consistent with this assumption.

In the past, we developed a sequence of work on interest point detection [22,19,23,20]. We first proposed an imbalance oriented selection scheme as an alternative of non-maximum suppression [22]. Next, we proposed a general formulation of imbalance [19]. Then, we proposed a global-to-local scheme [23] for imbalanced point matching. Recently, we proposed an association rule [20] to assign a scale to imbalanced points. In this paper, we propose a hybrid representation of imbalanced points for the two-layer matching scheme, based on the locality characteristics of imbalanced points and the tradeoff between the patch-type and SIFT-type representations. We use SIFT-type descriptors of imbalanced points in the first-layer matching (i.e., locality correspondence) in order to tolerate the inaccurate localization of interest points