Image and Vision Computing 32 (2014) 559–567

Contents lists available at ScienceDirect

Image and Vision Computing

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

Keypoint descriptor matching with context-based orientation estimation ☆

F. Bellavia [a,*], D. Tegolo [b], C. Valenti [b]

[a] CVG, Università degli Studi di Firenze, 50139 Firenze, Italy
[b] DMI, Università degli Studi di Palermo, 90123 Palermo, Italy

ARTICLE INFO

Article history:
Received 9 December 2013
Received in revised form 17 April 2014
Accepted 22 May 2014
Available online 2 June 2014

Keywords:
Image descriptors
Local features
Dominant orientation
Rotation invariance
Keypoint matching
SIFT
LIOP
MROGH

ABSTRACT

This paper presents a matching strategy to improve the discriminative power of histogram-based keypoint descriptors by constraining the range of allowable dominant orientations according to the context of the scene under observation. This can be done when the descriptor uses a circular grid and quantized orientation steps, by computing or providing a global reference orientation based on the feature matches.

The proposed matching strategy is compared with the standard approaches used with the SIFT and GLOH descriptors and the recent rotation invariant MROGH and LIOP descriptors. A new evaluation protocol based on an approximated overlap error is presented to provide an effective analysis in the case of non-planar scenes, thus extending the current state-of-the-art results.

1. Introduction

Keypoints extracted from digital images have been adopted with good results as primitive parts in many computer vision tasks, such as recognition [1], tracking [2] and 3D reconstruction [3]. The detection and extraction of meaningful image regions, named keypoints or image features, are usually the first step of these methodologies. Numerical vectors that embody the image region properties are successively computed to compare the keypoints found according to the particular task.

Different feature detectors have been proposed during the last decade invariant to affine transformations or scale and rotation only, including, but not limited to, corners and blobs. The reader may refer to [4] for a general overview.

After the keypoint is located, a meaningful descriptor vector to embody the characteristic properties of the keypoint support region (i.e. its neighborhood) is computed. Different descriptors have been developed, which can be divided mainly into two categories: distribution-based descriptors and banks of filters. In general, while the former give better results, the latter provides more compact descriptors. Banks of filters include complex filters, color moments, the local jet of the keypoint, differential operators and Haar wavelet coefficients. Refer to [5] for more details.

Distribution-based descriptors, also named histogram-based descriptors, divide the keypoint region, also called feature patch, into different areas and compute specific histograms related to some image properties for each area. The final descriptor is given by the ordered concatenation of these histograms. The rank and the census transforms [6], which consider binary comparisons of the intensity of central pixel against its neighborhood, are the precursors of the histogram-based descriptors. In particular, the CS-LBP [7] descriptor can be considered an extension of this kind of approach. The spin image descriptor, the shape context and the geometric blur and the more recent DAISY, BRIEF, BRISK and FREAK descriptors (see [5,8]) should be mentioned.

One of the most popular descriptors based on histograms is surely the SIFT (Scale Invariant Feature Transform) [9], which is a 3D histogram of gradient orientations on a Cartesian grid. SIFT has been extended in various ways since its first introduction. The PCA-SIFT descriptor [10] increases the robustness of the descriptor and decreases its length by applying PCA (Principal Component Analysis), RIFT (Rotation Invariant Feature Transform) [11] is a ring-based rotational invariant version, while GLOH (Gradient Local Orientation Histogram) [5] combines a log-polar grid with PCA and SURF [12] is an efficient discrete SIFT variant. Recently, RootSIFT [13] improves upon SIFT by replacing the Euclidean distance with the Bhattacharyya distance after the normalization of the descriptor vector with the Manhattan norm instead of the conventional Euclidean norm. Overlapping regions using multiple support regions combined by intensity order pooling are used by MROGH (Multi Support Region Order Based Gradient

☆ This paper has been recommended for acceptance by Cornelia M Fermuller.

* Corresponding author.
E-mail addresses: fabio.bellavia@unifi.it (F. Bellavia), domenico.tegolo@unipa.it (D. Tegolo), cesare.valenti@unipa.it (C. Valenti).

http://dx.doi.org/10.1016/j.imavis.2014.05.002
0262-8856/

Histogram) [14]. Furthermore, LIOP (Local Intensity Order Pattern) [15] uses the intensity order pooling and the relative order of neighbor pixels to define the histogram.

Over the last few years, machine learning techniques have been applied to remove the correlation between the descriptor elements and to reduce the dimension [10,16], as well as different histogram distances to improve the matches [17,18].

Different methodologies for evaluating feature descriptors and detectors have been proposed [4,5,8,16,19–23]. In the case of planar images, the Oxford dataset benchmark [4,24] is a well-established set of de facto standard, although an extension to non-planar images is not immediate [19]. Other evaluation methodologies use laser-scanner images [21] or structure from motion algorithms [16,23] or epipolar reprojection on more than two images [20], but in general they require a complex and error prone setup.

1.1. Our contributions

This paper presents in Section 2 a matching strategy to improve the discriminative power of histogram-based keypoint descriptors by constraining the range of allowable orientations according to the scene context.

We build the proposed matching strategy on the sGLOH (shifting GLOH) descriptor described in Section 2, presented in our previous work [25]. It uses a circular grid to embed more descriptor instances with different dominant discrete orientations of the same feature patch into a single feature vector. Each descriptor instance is accessible by an internal shift of the feature vector elements without the need to recompute the histograms. The matching distance between features is modified to consider the minimum distance among all descriptor instances for the possible dominant discrete orientations.

The sGLOH design can be used to further constrain the allowable dominant discrete orientations to be considered in the matching distance. A finer selection of the range of the dominant discrete orientations to be considered can be done a priori by defining a very fast matching strategy, named sCOr (shifting Constrained Orientation) or alternatively, when no further information is given in advance, by using an adaptive distance measure according to a voting strategy to get the sGOr (shifting Global Orientation) matching (see Section 2).

In order to assess the properties of the novel matching strategies, different experiments reported in Section 3 were carried out, both on planar and non-planar scenes. To provide more insights, we also evaluated the case when more than just the first dominant orientation is used in SIFT and GLOH. The rotational invariant MROGH [14] and LIOP [15] were also included in the evaluation due to the increasing interest towards them in recent works [14,8].

In the case of non-planar scenes, a novel dataset was created which employs a new evaluation protocol based on the approximated overlap error [26,27]. This evaluation protocol provides an effective analysis in the case of non-planar scenes, extending the current state-of-the-art results [20,22]. Section 4 reports final comments and conclusions.

2. The proposed matching strategy

Patch normalization and orientation methods are presented before defining the keypoint matching with sCOr and sGOr, as well as details on the sGLOH descriptor [25], which is essential in the matching pipeline since it allows constraints on the range of allowable orientations.

2.1. Patch normalization

Given an image $I(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2$, the feature patch must be normalized before the computation of the descriptor vector. In the general case of affine-invariant keypoint detectors, which usually represent a good trade-off between transformation invariance and discriminative power, an elliptical region $\mathcal{R} \subset I$ is extracted. If the ellipse is given by the equation $\mathbf{x}^T\Sigma^{-1}\mathbf{x} = 1$, considering the keypoint center as coordinate origin, the patch is normalized to a circle of a fixed radius r according to the formula $\mathbf{x}' = rA\mathbf{x}$, where $A = D^{-\frac{1}{2}}R^T$ with $\sum = RDR^T$ by the eigenvalue decomposition [5]. The ellipse axis lengths and orientations are given by the square root of the eigenvalues and the corresponding eigenvectors of Σ, respectively [5]. The symmetric matrix $\Sigma \in \mathbb{R}^{2\times 2}$ is obtained as the covariance matrix for some quantity $\phi(\mathbf{x}) \in \mathbb{R}^2$ related to the points of the patch $\mathcal{R}$ [4]. This is the gradient vector in the case of the Harris detector or the coordinates of the boundary points for the MSER detector [4].

The affine illumination invariance is obtained by normalizing the intensity value $I(\mathbf{x})$ of the points inside the region $\mathcal{R}$ through their mean μ and standard deviation σ, according to the formula $I\prime(\mathbf{x}) = \frac{I(\mathbf{x})-\mu}{\sigma}$.

2.2. Finding the patch orientation

In order to be rotational invariant, most of the histogram-based descriptors have to be rotated according to a reference dominant orientation and different methodologies have been designed for its computation [9,12,28–30].

The common approach was proposed by Lowe [9], where the gradient $\nabla I(\mathbf{x}) = [dx,dy]^{\mathrm{T}}$ is computed for each point $\mathbf{x} \in \mathcal{R}$ and a histogram of orientations $\theta_{\nabla I(\mathrm{x})} = \arctan(dy/dx)$ is built up. The contribution of each point $\mathbf{x}$ is given by its gradient magnitude $\|\nabla \boldsymbol{I}(\mathbf{x})\|$, weighted by the Gaussian function

$$g_\sigma(\mathbf{x}) = \frac{1}{2\pi\sigma^2} e^{-\frac{\|\mathbf{x}\|^2}{2\sigma^2}}$$

with standard deviation σ. It is assumed that $\theta_{\mathbf{v}} = \arctan(\mathbf{v})$, i.e. the arctangent of a generic vector $\mathrm{v} \in \mathbb{R}^2$, and the coordinate origins coincide with the center $\mathrm{x_c}$ of the feature patch. The orientation Θ of the highest peak in the histogram, interpolated by a parabolic fitting, is chosen as the dominant orientation. More dominant orientations can be assigned to the feature, by retaining other peaks above 80% of the highest peak [9].

Histogram-based descriptors which do not require a dominant orientation have also been proposed [15,14]. The underlying idea, first exploited by RIFT [11], consists in the use of the outside direction perpendicular to the tangent direction at each point as reference in the bin assignment. This is however not sufficient, because patch regions have to be made invariant to rotation, too. The RIFT descriptor uses concentric rings which are clearly rotational invariant, but less discriminative.

A last approach, named intensity order pooling [14], defines regions according the intensity value of the feature patch points without spatial constrains, i.e. points with similar intensity values belong to the same region. This approach is used by MROGH [14], which computes a gradient histogram as for RIFT, and by LIOP [15], where instead the bins represent the relative order of the intensities in a neighborhood of the point. Furthermore, MROGH uses multiple support regions, which result in overlapping regions. According to [8], both LIOP and MROGH seem to outperform recent state-of-the-art descriptors, at least in the case of the Oxford planar scenes. However, it must be noted that the MROGH outer support region is 2.5 times bigger than the standard elliptic region employed by other descriptors [14] in the tests, which lead to better but distorted results, especially in the case of planar scenes. See the additional material formore details.

The sGLOH approach [25] uses the RIFT reference orientation for bin assignment, but a circular grid is maintained instead of concentric rings. The descriptor instances for different dominant discrete orientations obtained by shifting the descriptor vector are compared during the