J. Vis. Commun. Image R. 20 (2009) 511–520

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Techniques for efficient and effective transformed image identification

Mohammad Awrangjeb [a,*], Guojun Lu [b,1]

[a] Cooperative Research Center for Spatial Information, The University of Melbourne, 723 Swanston St, Carlton, Vic. 3053, Australia
[b] Gippsland School of IT, Monash University, Churchill, Vic. 3842, Australia

ARTICLE INFO

Article history:
Received 26 March 2009
Accepted 15 July 2009
Available online 23 July 2009

Keywords:
Feature detection
Feature representation
Feature matching
Transformed image identification

ABSTRACT

In many applications, one common problem is to identify images which may have undergone unknown transformations. We define this problem as *transformed image identification* (TII), where the goal is to identify geometrically transformed and signal processed images for a given test image. The TII consists of three main stages – *feature detection*, *feature representation*, and *feature matching*. The TII approach by Lowe [D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision 60 (2) (2004) 91–110] is one of the most promising techniques. However, both of its feature detection and matching stages are expensive, because a large number of feature-points are detected in the image scale-space and each feature-point is described using a high dimensional vector. In this paper, we explore the use of different techniques in each of the three TII stages and propose a number of promising TII approaches by combining different techniques of the three stages. Our experimental results reveal that the proposed approaches not only improve the computational efficiency and decrease the storage requirement significantly, but also increase the transformed image identification accuracy (robustness).

1. Introduction

In many applications, such as image copyright protection [2], face recognition [3] and object recognition [1], a common problem is to identify images which may have undergone unknown transformations. We define this common problem as the *transformed image identification* (TII), where the goal is to identify the geometric transformed and the signal processed images for a given image. Therefore, the TII is different from conventional *content-based image retrieval* (CBIR) [4], where all images having the same or similar features, e.g., similar colors, are considered relevant to each other. Li et al. [5] proposed a user feedback-based CBIR technique using learning machines for classification [6].

The TII is also called *near-duplicate image identification* in the literature [3,7–10] and consists of three main stages [11]: *feature detection*, *feature representation*, and *feature matching*. In the *feature detection* stage, a set of features, e.g., corners, blobs, T-junctions, are detected. The most valuable property of a feature detector is *repeatability*, i.e., whether it reliably finds the same feature-points after the image has undergone different transformations. In the *feature representation* stage, each detected feature-point is represented by a feature vector calculated possibly from its neighborhood. In the *feature matching* stage, the feature vectors of the test image and the stored images are compared to identify transformed images for the test image. The matching is often based on a distance, e.g., the Euclidean distance [11], between the feature vectors.

Low's approach [1] has demonstrated its superior performance in identifying transformed images over many other approaches [12–14]. However, both of its detection and matching stages are expensive, because a large number of keypoints are detected in the scale-space using the difference-of-Gaussian (DoG) filter and each keypoint is described using a 128-dimensional vector known as SIFT (scale invariant feature transform) descriptor. Further research has been carried out in the literature for the dimensionality reduction of the SIFT descriptor, but with the expense of loss of robustness [11–13]. In this paper, we explore the use of different techniques in each of the TII stages. In stage one, we present two possible solutions for feature-point reduction. First is to down scale the image before the DoG keypoint detection and second is to use corners (instead of DoG keypoints) which are visually significant, more robust, and much smaller in number than the DoG keypoints. In stage two, we explore the use of corner–curvature as well as to SIFT descriptors. In stage three, feature-point matching techniques based on the geometric point matching technique [15] is used in addition to the nearest-neighbor-distance-ratio based matching technique [1]. Consequently, we propose a number of TII approaches in this paper by combining different techniques at different stages and will discuss them in Section 4. The two feature-point reduction solutions combined with the SIFT descriptors and our previously proposed feature-point matching technique

* Corresponding author. Fax: +61 3 9349 5185.
E-mail addresses: mawr@unimelb.edu.au (M. Awrangjeb), Guojun.Lu@infotech.monash.edu.au (G. Lu).
[1] Fax: +61 3 5122 6879.

1047-3203/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jvcir.2009.07.004

[15] not only improve the computational efficiency and decrease the storage requirement significantly, but also increase the transformed image identification accuracy (robustness).

The organization of this paper is as follows. In Section 2, we briefly review the existing feature detection, representation, and matching techniques and highlight key contributions of the paper. Section 3 compares three existing feature detectors in terms of robustness and detection time. In Section 4, we present the proposed TII approaches. In Section 5, we discuss the performance study and finally, in Section 6 we conclude the paper.

2. Related work and contributions of this paper

In this section, we first briefly present some relevant existing work in three stages of the TII (feature-point detection, feature representation and feature matching). Then we briefly discuss existing promising TII approaches and present the contribution of this paper.

2.1. Feature detection

A large number of corner and interest-point detectors have been proposed in the literature [1,16,14,17–23]. While corner detectors detect image spatial locations where edge segments make significant angles, interest-point detectors not only detect corners, but also image locations that have large gradients in all directions at a predetermined scale [1].

All corner and interest-point detectors can be broadly classified into two groups: single-scale detectors [21–23] and multi-scale detectors [1,14,16–20]. Single-scale detectors work well if the image has similar size features, but ineffective otherwise; because either fine or coarse scale feature is poorly detected, but images may contain both kinds of features. To improve the effectiveness of the detection stage, multi-scale detectors have been proposed.

Corner and interest-point detectors can also be categorized into three groups: intensity-based [1,16,14,22], contour-based [17–21], and model or template-based [23] methods. Intensity-based methods estimate a measure which is intended to indicate the presence of an interest-point directly from the image pixel values. Contour-based methods first obtain planar curves using some edge detector and then search for the curvature maxima along those curves. Model or template-based methods find corners by fitting the image signal into a predefined model.

The main drawback with the model-based detectors is that the corner in natural images cannot be approximated by a model of a perfect corner, as it can take any form of the bidirectional signal change [14]. Moreover, they are computationally too expensive [23] and are not used for general purpose [24]. This paper will focus on the intensity and contour-based detectors only.

2.1.1. Intensity-based detectors

Probably the most widely used detector is the Harris interest-point detector [22] which is based on the eigen values of the second-moment matrix. However, Harris points are not scale-invariant [11]. Lindeberg [16] introduced the concept of automatic scale selection which allows detecting interest points in an image, each with their own characteristic scale. Mikolajczyk and Schmid [14] refined this technique by creating robust and scale invariant features with high repeatability. They used a scale-adapted Harris measure or the determinant of the Hessian matrix to select the location, and the Laplacian to select the scale. Lowe's [1] approximation of the Laplacian of Gaussian using the DoG filter speeded up the feature detection stage significantly. The recently proposed fast-Hessian detector in the SURF (speeded up robust features) detector–descriptor scheme [11] used a basic approximation of the Hessian matrix and relied on the integral images to reduce the computational cost.

2.1.2. Contour-based detectors

The CSS (curvature scale-space) detector in [19] is one of the earlier contour-based multi-scale detectors. It detected corners at a high scale and tracked them through multiple lower scales in order to improve localization. Since different curves require different smoothing-scales and there may be different sizes of corners on the same curve, this detector is highly sensitive to the use of a single corner detection scale and a single fixed curvature–threshold. He and Yung [21] improved this detector by using the adaptive curvature–threshold and the dynamic region-of-support on both sides of each curvature extremum point. Zhang et al. [20] further improved it by introducing the idea of curvature product. In terms of the curvature product, the strong corners become more distinguishable from the weak and false corners. Awrangjeb et al. [17] proposed another improvement (known as affine resilient CSS – ARCSS – detector [25]) by selecting three corner detection scales based on the curve's affine-length.

Recently Awrangjeb and Lu [18] pointed out the two main problems of the above CSS-based detectors. First, the CSS curvature estimation technique is highly sensitive to the local variation and noise on the curve. Second, the CSS corner detection technique requires appropriate Gaussian smoothing-scale selection which is a difficult task. To overcome these two problems, they proposed a new corner detector [18] based on the *chord-to-point distance accumulation* (CPDA) for the discrete curvature estimation [26], which is less sensitive to the local variation and noise on the curve and does not require appropriate Gaussian smoothing-scale selection.

2.2. Feature representation

In order to facilitate feature-point matching in any subsequent application, each feature-point must be represented with some of its associated information. The more the representation is distinctive, the less the number of false candidate matches will be obtained in the matching stage. The feature representation is also known as the feature descriptor in the literature.

There are two different types of representations found in the literature: geometric descriptors and local descriptors.

2.2.1. Geometric descriptors

This type of representation [27,28] is purely geometric, where each corner or feature-point is represented using its curvature, angle, and distances from neighbor corner-points. Zhou et al. [27] used the angles of Delaunay triangles which are formed among the Harris interest-points [22]. These angles are invariant to image translation, rotation, and uniform scaling. Huttenlocher used [28] distance ratios defined by the quadruples of the feature-points. The distance ratios are invariant to affine transformations.

Awrangjeb and Lu [15] used the curvature descriptors for their proposed *geometric point matching* (GPM) technique. In fact, their proposed ARCSS [25] and CPDA [18] detectors provide various information available for later use. For each corner, these detectors provide its position, absolute curvature value, angle with its two neighboring corners or with the endpoints when the necessary number of neighbor corners are not found and the affine-lengths between neighboring corners on the same curve. All of the above information can be used collectively as the 'curvature descriptor' to represent the corner.

Though the geometric representation is easy to design and requires quite small amount of storage per feature-point, the representation is not unique. As a consequence, either the true correspondences are missed or a huge number of false correspondences are obtained between two images and the matching procedure