J. Vis. Commun. Image R. 38 (2016) 602–613

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Matching of images with projective distortion using transform invariant low-rank textures ☆

Qiang Zhang [a,b,*], Yajun Li [b], Rick S. Blum [c], Peng Xiang [b]

[a] Key Laboratory of Electronic Equipment Structure Design (Xidian University), Ministry of Education, Xi'an, Shaanxi 710071, China
[b] Center for Complex Systems, School of Mechano-Electronic Engineering, Xidian University, Xi'an, Shaanxi 710071, China
[c] Electrical and Computer Engineering Department, Lehigh University, Bethlehem, PA 18015, United States

ARTICLE INFO

Article history:
Received 24 November 2015
Revised 1 April 2016
Accepted 7 April 2016
Available online 8 April 2016

Keywords:
Image matching
Projective distortion
Transform invariant low-rank textures (TILT)
Automatic low-rank texture region detection
Geometric shape descriptor

ABSTRACT

A matching method is presented for images with projective distortion based on transform invariant low-rank textures (TILT). In the method, the problem of matching images with projective distortion is first reduced to a problem of matching two rectified images just with scaling and translation distortions via TILT. Then a point-feature based matching method is employed to establish the corresponding points between the two rectified images. This is different from some traditional methods that try to directly seek local affine or projective invariants from input images. Moreover, no prior knowledge on the epipolar geometry is required in the proposed method. An automatic low-rank texture region detection method is presented to make the method more applicable in practice. Additionally, a new descriptor is constructed by combining a proposed geometric shape descriptor and the traditional SIFT descriptor to further improve the correct matching rate. Experimental results demonstrate the validity of the proposed method.

1. Introduction

Image matching is an approach to identify the corresponding points in two images of the same scene taken at different times, from different sensors or from different viewpoints [1]. It has been widely applied in many computer vision and pattern recognition tasks, including object recognition [2,3], image stitching [4], broadcast video analysis [5,6], and 3D reconstruction [7–9]. Image matching methods have been previously developed [10–16]. They are very effective at finding matches in images with a limited number of distortions, such as similarity or affinity. However, real images often include more general distortions. In fact, real images with distortion that is well approximated by projective distortion are common.

If we restrict ourselves to planar scenes, the projective distortion between two images could be described by a plane projective transformation [17,18], i.e.,

$$\mathbf{p}' = \begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = H\mathbf{p}, \tag{1}$$

where the homography matrix H is a 3×3 non-singular matrix. In (1), $\mathbf{p} = [x, y, 1]^T$ denotes the homogenous coordinates of a point in the first image with coordinates $(x, y)^T$, while $\mathbf{p}' = [x', y', w']^T$ denotes the homogenous coordinates of the corresponding point in the second image with coordinates

$$\begin{bmatrix} \frac{x'}{w'} \\ \frac{y'}{w'} \end{bmatrix} = \begin{bmatrix} \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}} \\ \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}} \end{bmatrix}. \tag{2}$$

As discussed in [17], the projective transformation is a general non-singular linear transformation of homogeneous coordinates and combines the affine transformation with projectivity. Accordingly, the problem of matching images with projective distortion is more challenging than that of matching images with similarity or affine transformation.

Some approaches to the matching of images with projective distortions have been published in recent works, among which the local-invariant based methods are one of the most popular categories. Furthermore, one of such methods is based on the assumption that the projective transformation can be locally

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author at: P.O. Box 183, Department of Automatic Control, Xidian University, No. 2 South TaiBai Road, Xi'an, Shaanxi Province 710071, China.
E-mail address: qzhang@xidian.edu.cn (Q. Zhang).

http://dx.doi.org/10.1016/j.jvcir.2016.04.007
1047-3203/

approximated by a similarity or affine transformation. In these cases, some scale or affine invariants could be applied, such as the Scale Invariant Feature Transform (SIFT) [19], the Affine-SIFT (ASIFT) [20], and the method of Maximally Stable Extremal Regions (MSER) [21]. However, in some cases the projective transformation cannot be approximated by a similarity or affine transformation. Some projective invariants have also been proposed to overcome this problem, which are defined on point sets [18,22], on sets composed both from points and straight lines [23,24] or on a shape basis [25].

In addition to those using local invariants for image matching, some methods were presented which employed geometric constraints among correspondences [26–29]. These methods first defined a proper measure of the geometric consistency among point correspondences, and then formulated the image matching problem as an optimization problem [23]. Usually, these methods achieve higher performance, but also have higher computational complexity. Moreover, an accurate estimation of the epipolar geometry is usually required in these methods [26,27], either through prior knowledge of the configuration or through a few strong matches to estimate the fundamental matrix.

In the past few years, sparse representation and low-rank representation [30–32] have attracted significant attention in the field of computer vision and image processing. More especially, Zhang et al. [33] proposed a new tool, transform invariant low-rank textures (TILT), based on sparse and low-rank matrix decomposition, to recover the "intrinsic" low-rank textures in the 3D scene from the deformed 2D images. Furthermore, it can remove the deformation caused by affine or projective transformation to produce a clean image captured from a front-viewpoint. Their experiments demonstrated that TILT could effectively and robustly work for a wide range of regular and near-regular patterns or objects in real images. So far, it has been successfully applied to many computer vision tasks, such as text extraction [34], camera calibration [35] and image rectification [36].

In this paper, we will present a method to address the matching of "low-rank texture images[1]" with projective distortion by using TILT. Similar to those in [26,27], we first rectify the reference image and the image to be matched, respectively, to reduce the projective distortion between the two original input images to some extent. Then we exploit a point-feature based matching method on the two rectified images to obtain the final corresponding points. We achieve the rectification of the input images via TILT, instead of employing any prior knowledge on the epipolar geometry as in [26,27]. Experimental results demonstrate that the proposed method performs better than some existing methods such as SIFT, ASIFT and MSRE, especially in cases with images experiencing severe projective distortions.

The main contributions of this paper are as follows: (1) A matching method for images with projective distortion is proposed based on TILT. Different from those traditional methods that try to directly seek local affine or projective invariant features from the input images, the proposed method reduces the problem of matching images with projective distortion to a problem of matching images with translation and scale distortions via TILT. Moreover, it requires no prior knowledge on the epipolar geometry as in [26,27]. (2) An automatic low-rank texture region detection method is presented. With the proposed texture region detection method, the low-rank textures will be automatically, rather than manually as in [33], selected from the original input images before they are fed into TILT. (3) A novel descriptor is introduced for each considered point when the two rectified images are matched. In addition to the local information from each considered point, the geometric shape information of the pixels around each considered point is employed in the proposed descriptor.

The rest of the paper is organized as follows. Section 2 gives a brief introduction to TILT. In Section 3, the proposed image matching method is described in detail. Experimental results and some conclusions are given in Sections 4 and 5, respectively.

2. Transform invariant low-rank texture (TILT)

To improve the readability of this paper, the key idea of TILT, presented in [33], will be reviewed briefly in this section.

Considering a 2D texture image I_0, represented by a matrix (also denoted by I_0 for convenience) of order $m_1 \times m_2$, it is seen as a low-rank texture if the rank of the matrix I_0 is far less than its smaller dimension (m_1 or m_2) [33]. As discussed in [33], although the planar surfaces or structures in the 3D scene exhibit low-rank textures, their images do not necessarily have low rank because of deforming transformations and corruptions.

Given a deformed and corrupted image I of a low-rank texture

$$I = (I_0 + E) \circ H^{-1}, \tag{3}$$

the goal of TILT is to recover the "intrinsic" low-rank texture I_0 and the transformation $H : R^2 \to R^2$. In (3), E denotes the corruptions and is assumed to be sparse [33]. And $I \circ H$ denotes the transformed version of the image I using the transformation matrix H.[2] Mathematically, the problem of TILT is formulated as

$$\min_{I_0,E,H} rank(I_0) + \gamma\|E\|_0 \quad s.t.\ I \circ H = I_0 + E, \tag{4}$$

where $\|E\|_0$ denotes the l_0-norm of the matrix E (i.e., the number of the non-zero entries in the matrix E), and $\gamma > 0$ is a weighting parameter that trades off the importance of the rank of the texture and the sparsity of the error.

The non-convex rank and l_0-norm minimization problems in (4) can be, respectively, relaxed by the convex nuclear-norm and l_1-norm minimization [33,37]. The nonlinear constraint function $I \circ H = I_0 + E$ can also be linearized as $I \circ H + \nabla I \Delta H = I_0 + E$, where ∇I is the Jacobian of the image I with respect to the transformation parameters.[3] Thus the problem in (4) reduces to the following convex optimization problem

$$\min_{I_0,E,H} \|I_0\|_* + \gamma\|E\|_1 \quad s.t.\ I \circ H + \nabla I \Delta H = I_0 + E, \tag{5}$$

where $\|\cdot\|_*$ denotes the nuclear-norm of a matrix and is defined as the sum of its singular values. $\|\cdot\|_1$ denotes the l_1-norm of a matrix and is defined as the sum of the absolute values of its entries. The optimization problem in (5) can be effectively solved by using an iterative linearization scheme. More details are provided in [33].

Fig. 1 illustrates some results of TILT. As shown in Fig. 1, the ranks of the image matrices for the recovered textures are less than those of the image matrices for the deformed images. In addition to the recovered low-rank textures, the local distortion transformations between the deformed low-rank textures and their correspondingly recovered ones can also be obtained by using TILT. As shown in Fig. 1(d)–(f), the deformed low-rank textures will be approximately rectified to some new ones by these local transformations. Accordingly, a coarsely rectified image will also be obtained when the local transformation is applied to the whole

[1] As discussed in [33], the low-rank texture images generally denote those containing large a number of regular or symmetrical structures.

[2] For example, if H denotes a projective transformation in (1), then $(I \circ H)(x, y)$ may be computed as $I\left(\frac{h_{11}x+h_{12}y+h_{13}}{h_{31}x+h_{32}y+h_{33}}, \frac{h_{21}x+h_{22}y+h_{23}}{h_{31}x+h_{32}y+h_{33}}\right)$.

[3] Strictly speaking, as discussed in [33], ∇I is a 3D tensor: it gives a vector of derivatives at each pixel whose length is the number of the parameters in the transformation H. When ∇I is multiplied by another matrix or vector, it contracts in the obvious way which should be clear from the context.