Rotational Invariants for Wide-baseline Stereo *

Jiří Matas, Petr Bílek, Ondřej Chum

Centre for Machine Perception

Czech Technical University, Department of Cybernetics

Karlovo namesti 13, Prague, Czech Republic

tel: +420 2 2435 7637, fax: +420 2 2435 7385

e-mail: [matas,bilek,chum]@cmp.felk.cvut.cz

Abstract

The problem of establishing correspondences between a pair of images taken from different viewpoints, i.e. the "wide-baseline stereo" problem, is studied in the paper. To handle the problem of affine distortion between two corresponding regions a method based on rotational invariants computed on normalised measurement regions is applied.

A robust similarity measure for establishing tentative correspondences is used. The robustness ensures that invariants from multiple measurement regions, some that are significantly larger (and hence discriminative) than the distinguished region, may be used to establish tentative correspondences.

1 Introduction

Given two images of a scene taken from arbitrary viewpoints, the problem of establishing reliable correspondences is fundamental in many computer vision tasks. Applications include 3D scene reconstruction, motion recovery, image mosaicing, content-based image retrieval, mobile robot navigation and many more. In the wide-baseline set-up, local image deformation cannot be realistically approximated by translation or translation with rotation, and a full affine model is required. Correspondence cannot be therefore established by comparing regions of a fixed shape, like rectangles or circles, since their shape is not preserved under the group of transformations that occur between the images.

In the literature, correspondences have been traditionally sought by matching features computed on local neighborhoods of detected interest points [18, 10, 1, 12]. To cope with different viewpoint, both the local regions and the descriptors of such regions have to be defined in affine

*This research was supported by the Czech Ministry of Education under Research Programme MSM 210000012 Transdisciplinary Biomedical Engineering Research and Grant Agency of the Czech Republic GACR 102/01/0971.

invariant way. The fully affine-invariant regions were introduced recently, exploiting local texture characteristics [1], or local configuration of multiple image edges or interest points [5, 16]. Schaffalitzky and Zisserman [7] presented a method for automatic determination of local neighborhood shape, but only for image areas where stationary texture occurs.

In this paper, we rely on the so called Maximum Stable Extremal Regions and Separated Elementary Cycles of the Edge Graph introduced in [4], which were shown to define highly repeatable local frames over a wide range of image formation conditions. Using measurements on these frames, we are able to successfully solve non-trivial instances of the problem of establishing correspondences between two images. We experimentally show that the measurements are sufficiently stable.

The main contribution of the paper is the utilization of processes for determination of fully affine-invariant descriptors of local regions. The approach is based on moment invariants. However, instead of using full affine invariants [15, 2], we first normalise local region up to rotation and then only the rotational invariants are computed.

The paper is organised as follows: The structure of the class of wide-baseline and recognition algorithms and two types of distinguished regions (originally proposed by Matas et al. [4]) are discussed in Section 2 and Section 3. Section 4 aims to main contribution of this paper, i.e. identifying measurement regions and extracting their affine invariant characterisations.

In Section 5 details of a matching algorithm (from the above-mentioned class) are given. A *robust* approach proposed by Matas et al. [4] is used for tentative correspondence computation.

Experimental results on images taken with an uncalibrated camera are presented in Section 6. Epipolar geometry is established using combination of multiple types of distinguished regions. Presented experiments are summarised and the contributions of the paper are reviewed in Section 7.

2 Correspondence from Distinguished Regions

Algorithms for wide-baseline stereo described in the literature have adopted strategies with a similar structure whose core can summarised by concept based on distinguished regions (introduced by Matas et al. [4]):

Algorithm 1: Wide-baseline Stereo from Distinguished Regions - The Framework

1. Detect **distinguished regions**.
2. Describe DRs with invariants computed on **measurement regions**.
3. Establish **tentative correspondences** of DRs.
4. **Estimate epipolar geometry** in a hypothesise-verify loop.

Distinguished Regions. To identify correspondences between two images, simply detectable and stable regions have to be present in the images. We will call such regions **distinguished regions** (DR). Matas et al. [4] have defined DR in more formal way:

> Let image I be a mapping $I : \mathcal{D} \subset \mathbb{Z}^2 \to \mathcal{S}$. Let $\mathcal{P} \subset 2^{\mathcal{D}}$, i.e. $\mathcal{P}$ is a subset of the power set (set of all subsets) of $\mathcal{D}$. Let $\mathcal{A} \subset \mathcal{P} \times \mathcal{P}$ be an adjacency relation on $\mathcal{P}$ and let $f : \mathcal{P} \to \mathcal{T}$ be any function defined on $\mathcal{P}$ with a totally ordered range $\mathcal{T}$. A **region** $\mathcal{Q} \in \mathcal{P}$ **is distinguished** with respect to function f iff $f(\mathcal{Q}) > f(\mathcal{Q}'), \forall(\mathcal{Q}, \mathcal{Q}') \in \mathcal{A}$.

Measurement Regions. Note that we do not require DRs to have any transformation-invariant property that is unique or rare in the image. In other words, DRs need not be discriminative (salient). If a local frame of reference is defined on a DR by a transformation-invariant construction (projective, affine, similarity invariant), a DR may be characterised by invariant measurements computed on any part of an image specified in the local (DR-centric) frame of reference. We used the term **measurement region** for this part of the image.

Invariant Descriptors. The most simple situation arises if a local affine frame is defined on the DR. Photometrically normalised pixel values from a normalised patch characterise the DR invariantly. More commonly, only a point or a point and a scale factor are known, and rotation invariants [9, 8] or affine invariants [15, 2] must be used.

Tentative Correspondences. At this stage, we have a set of DRs for each image and a potentially large number of invariant descriptors associated with each DR. Selecting mutually nearest pairs in Mahalanobis distance is the most common method [8, 15, 9]. Note that the objective of this stage is not to keep the maximum possible number of good correspondences, but rather to maximise the fraction of good correspondences. The fraction determines the speed of epipolar geometry estimation by the RANSAC procedure [13].

Epipolar Geometry estimation is carried out by a robust statistical method, most commonly RANSAC. In RANSAC, randomly selected subsets of tentative correspondences instantiate an epipolar geometry model. The number of correspondences consistent with the model defines its quality. The hypothesise–verify loop is terminated when the likelihood of finding a better model falls below a predefined threshold.

3 Detection of DR

The art is in finding distinguishing properties that can be detected without the obviously prohibitive exhaustive enumeration of all subsets. We employ new types of distinguished regions proposed by Matas et al. [4]. For both types of DR, *Separated Elementary Cycles of the Edge Graph (SECs)* and the *Maximally Stable Extremal Regions (MSERs)*, an efficient (near linear complexity) and practically fast (from fraction of a second to seconds) detection algorithm has been found. Low computational complexity and invariance to photometric and geometric transformation are desirable theoretical properties of the process of distinguished region detection. Stability, robustness and frequency of detection and hence usefulness of a particular type of DR has been tested experimentally and successful wide-baseline experiments on indoor and outdoor datasets was presented.

4 Affine Invariant Description of DR

4.1 Affine Invariant Measurement Region

If we have identified DR, we would like to characterize it by measurements computed on part of an image (measurement region, MR) defined by this DR. In order to cope with different viewing conditions the MRs and descriptors extracted from MR have to be defined in an invariant way. We will assume that only affine transformation is present.

To define measurement region we use first μ_1 and second Σ_1 statistics computed on data positions within distinguished region. The mean μ_1 and covariance matrix Σ_1 define ellipse $E_1(x, y)$:

$$(\mathbf{x} - \mu_1)^T \Sigma_1^{-1} (\mathbf{x} - \mu_1) = 1, \tag{1}$$

where $\mathbf{x} = (x, y)^T$. Without loss of generality we will further assume $\mu_1 = \mu_2 = (0, 0)$. First, we will prove that covariance matrix Σ_1 and covariance matrix Σ_2 computed on original DR and DR after affine transformation A are also related by A. Let:

$$\Sigma_1 = \frac{1}{|\Omega_1|} \int_{\Omega_1} \mathbf{x}\mathbf{x}^T \, d\Omega_1 \quad \text{and} \quad \mathbf{y} = A\mathbf{x}, \tag{2}$$

then

$$\Sigma_2 \;=\; \frac{1}{|\Omega_2|} \int_{\Omega_2} \mathbf{y}\mathbf{y}^T \, d\Omega_2 = \frac{1}{|A||\Omega_1|} \int_{\Omega_1} A\,\mathbf{x}\mathbf{x}^T A^T |A| \, d\Omega_1 = A\,\Sigma_1\,A^T, \tag{3}$$

where Ω_1 and Ω_2 are regions defined by first and second distinguished region.

Next we will prove that if the distinguished region is transformed by affine transform A, then the transformation between ellipses E_1 and ellipse E_2 (defined by Σ_2 computed from transformed region) is known:

$$\mathbf{y}^T \Sigma_2^{-1} \mathbf{y} = (A\mathbf{x})^T (A\Sigma_1 A^T)^{-1} (A\mathbf{x}) = \mathbf{x}^T A^T A^{-T} \Sigma_1^{-1} A^{-1} A\mathbf{x} = \mathbf{x}^T \Sigma_1^{-1} \mathbf{x}. \tag{4}$$

It means, both the original and the transformed ellipses are related by affine transformation A. The E_1 can be transformed to E_2 (and vice versa), nonetheless, this transformation is known up to rotation. Indeed, if we denote $\Sigma_2 = C^T C$, we can write for arbitrary rotation R:

$$\mathbf{y}^T \Sigma_2^{-1} \mathbf{y} = \mathbf{y}^T (C^T C)^{-1} \mathbf{y} = \mathbf{y}^T (C^T \underbrace{R^T R}_{E} C)^{-1} \mathbf{y}. \tag{5}$$

The ellipse defined by covariance matrix will be used for MR normalisation up to rotation in next step.

4.2 Affine Invariant Description

There exist three basic ways how to obtain characterisation of DR that will be invariant to affine transformation: 1. compute affine invariant directly from MR, 2. identify local affine coordinate system and normalise MR or 3. normalise MR up to rotation and compute rotation invariant.

The first two approaches are used in many recent image matching and wide-baseline stereo algorithms [6, 15, 14]. In this paper we are focused on third approach, i.e. we normalise all MRs up to rotation (exploiting results from previous paragraph) and on this region compute characterisation that is invariant to rotation (see Fig. 1). It is obvious that this characterisation has to be also affine invariant.

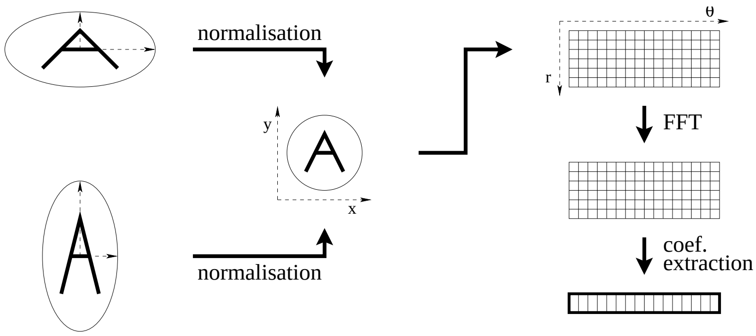

Figure 1: Computing affine invariant descriptors

We have to find a transformation which turns ellipse defined by covariance matrix to unity circle (whitening of covariance matrix). By this transformation we 'normalise' measurement region. Since corresponding MRs are same up to an affine transform, the normalised MR are same up to an unknown rotation. Inn order to match corresponding MR we have to determine this rotation.

The requirement that the rotation have to be known can be relaxed by employing rotation invariants. We exploit invariants based on integral transform (moment invariants). Let $I(x, y)$ be normalised MR, then rotational moment of order $k + l$ is defined as

$$M(k, l) = \iint P_k(r) e^{-i\theta l} I(r, \theta)\, d\theta dr, \tag{6}$$

where $I(r, \theta) = I(r \cos\theta, r \sin\theta)$ and $P_k(r)$ is polynomial with degree k. Rotational invariant descriptor is then computed as magnitude of moment with given order:

$$R_{inv}(k, l) = |M(k, l)|. \tag{7}$$

In our experiments we use $P_k(r) = r^k$, $k = 0, 1, 2$. Note, the other types of moments (Zernike, Fourier-Mellin, Complex etc.) are special cases [17, 11] for particular choice of function $P_k(r)$.

The algorithm for extracting affine invariant characterisation of DR is simply deduced from equations and can be summarised as follows (see also Fig. 1):

Algorithm 2: Extracting affine invariant characterisation of DR

1. compute first and second order statistics of MR

2. transform MR to have unit covariance matrix

3. express data of normalised MR in polar coordinates,
4. apply one-dimensional FFT along θ-axis and keep only magnitude of complex numbers,
5. combine coefficients along r-axis according to polynomial $P_k(r)$.

---

5 Matching

For establishing tentative correspondence we employ robust matching method proposed by Matas et al. [4]. Each DR is described by a measurement vector $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$. In the matching problem there are two sets $\mathcal{L}$ and $\mathcal{R}$ of DR measurement vectors originating from the 'left' and 'right" image respectively. The task is to find tentative matches given the local description. The set of initial correspondences is formed as follows. Two regions with descriptions $\mathbf{x} \in \mathcal{L}$ and $\mathbf{y} \in \mathcal{R}$ are taken as a candidates for a match iff $\mathbf{x}$ is the most similar measurement to $\mathbf{y}$ and *vice-versa*, i.e.

$$\forall \mathbf{x}' \in \mathcal{L} \setminus \mathbf{x} : d(\mathbf{x}, \mathbf{y}) < d(\mathbf{x}', \mathbf{y}) \qquad \text{and} \qquad \forall \mathbf{y}' \in \mathcal{R} \setminus \mathbf{y} : d(\mathbf{y}, \mathbf{x}) < d(\mathbf{y}', \mathbf{x}),$$

where d is the asymmetric similarity measures defined below. In the computation of $d(\mathbf{x}, \mathbf{y})$ each component of the measurement vector is treated independently. The similarity between the i-th component of $\mathbf{x}$ and $\mathbf{y}$ is measured by the number of vectors $\mathbf{y}'$ whose i-th measurement is closer. In other words the similarity in the i-th component is the rank of the measurement from $\mathbf{y}$ among all measurements $\mathbf{y}'$ from $\mathcal{R}$:

$$\text{rank}^i_{\mathbf{x},\mathbf{y}} = |\{A \in \mathcal{L} : |a_i - y_i| \leq |x_i - y_i|\}|.$$

The overall similarity measure is then defined as follows

$$d(\mathbf{x}, \mathbf{y}) = |\{i \in \{1, \dots, n\} : \text{rank}^i_{\mathbf{x},\mathbf{y}} < t\}|,$$

where n is dimension of the measurements vector and t a predefined ranking threshold. The computation of $d(\mathbf{y}, \mathbf{x})$ is analogous with the roles of $\mathcal{L}$ and $\mathcal{R}$ interchanged. The most important property of d is that the influence of any single measurement is limited to 1. Only the main idea of the probabilistic error model behind the design may be mentioned due to limited space. Under a very broad range of error models, corresponding measurements are more likely to be below the ranking threshold than a mismatch.

6 Experiments

The following parameters of the matching algorithm were used in following experiments. As distinguished regions we have tacitly used combination of **Separated Elementary Cycles of the Edge Graph** (SECs) and **Maximally Stable Extremal Region** (MSERs). These DR was detected by method proposed by Matas et al. [4]. The measurement regions defined in terms of affine-invariant constructions on the DR boundaries were the following: the DR itself and its convex hull scaled by factors of 1.5, 2 and 3. The MRs were described by affine invariant characterization as proposed in Section 4. Tentative correspondences comprised only those pairs whose characterisation were mutually nearest in the robust similarity measure. Epipolar geometry was estimated by the 7-point algorithm [3]. In all experiments, only a linear algorithm is used [3] to estimate epipolar geometry; no effort was made to improve the precision by known methods such as bundle adjustment, correlation, or homography growing.

6.1 Experiment I: Stability of measurement regions

The stability of assignment of measurement regions to distinguished regions is experimentally validated. In Fig. 2 there are examples of two corresponding DR with MR (fitted ellipse) in the left and right column. The MR are normalised and expressed in polar coordinates (r is vertical and θ horizontal axis). The result is depicted in the middle column. It is obvious, the normalised MR are same up to translation in r (horizontal) axis, which corresponds to rotation in Cartesian coordinates.

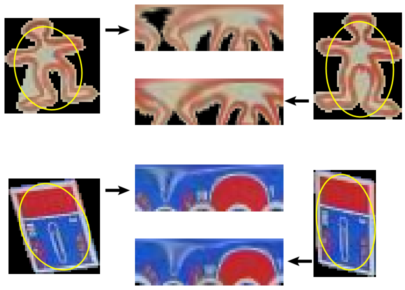

Figure 2: Stability of MR (see text for comments).

6.2 Experiment II: Epipolar geometry estimation

We tried to estimate epipolar geometry on the image from BOOKSHELF. The number of DRs in the left and right images was 1091 and 1118 respectively. The number of DRs with mutually nearest invariant descriptions, i.e. the number of tentative correspondences, was 424 in this test.

The RANSAC procedure found an epipolar geometry consistent with 187 tentative correspondences of which all are correct. The numbers of detected DRs in the left and right images, tentative correspondences (TC), epipolar geometry consistent correspondences (EG) and the number of mismatches (miss) are summarised in the caption of Figure 3. Mismatches are correspondences consistent with the estimated epipolar geometry that are not projections of the same part of the scene. The ratio TC/EG determines the average number of RANSAC hypothesis-verify attempts and hence the speed of epipolar geometry estimation.

The bottom row of Figure 3 shows close-ups of two rectangular regions selected from the left and right images respectively.

In the experiment conducted by Matas et al. [4] (the same dataset, but DR described by affine invariants) the number of tentative correspondences was 52, from which 29 was consistent. It is obvious that both methods provide comparable results, but further experiments on a wide range of scenes are needed to improve understanding of their relative merits.

Figure 3: BOOKSHELF: Estimated epipolar geometry (top row) and a close-up of selected areas marked with black rectangles in the originals (bottom row).

	left	right	TC	EG	miss
DRs	1091	1118	424	187	0

7 Conclusions

The main contribution of the paper is the method for defining affine invariant measurement regions and manner how the invariant characterisation of MRs is computed. We establish local affine frame, that is determined up to rotation and such rotation is eliminated by employing rotational invariants. This approach is similar to computation of affine invariants, nonetheless it can provide better stability. Moreover, it can serve as 'another detector of correspondences' in existing application and improve estimation of epipolar geometry.

In a second contribution, a robust similarity measure for establishing tentative correspondences was used. Due to the robustness, we were able to consider invariants from multiple measurement regions, even some that were significantly larger (and hence probably discriminative) than the associated distinguished region.

Good estimates of epipolar geometry were obtained on wide-baseline problems with the robustified matching algorithm operating on the output produced by the proposed detectors of distinguished regions. Fully affine distortions was present in the tests. Nonetheless, further experiments on a wide range of scenes are needed to improve understanding of merits of rotational and affine invariants.

References

[1] A. Baumberg. Reliable feature matching across widely separated views. In *Proc. of Computer Vision and Pattern Recognition*, pages I:774–781, 2000.

[2] J. Flusser and T. Suk. Pattern recognition by affine moment invariants. *Journal of the Pattern Recognition*, 26(1):167–174, 1993.

[3] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2000.

[4] Jiří Matas, Ondřej Chum, Urban Martin, and Tomáš Pajdla. Distinguished regions for wide-baseline stereo. Research Report CTU–CMP–2001–33, Center for Machine Perception, K333 FEE Czech Technical University, Prague, Czech Republic, November 2001.

[5] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *ICCV'98*, pages 754–760, 1998.

[6] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *Proc. 6th International Conference on Computer Vision, Bombay, India*, pages 754–760, January 1998.

[7] F. Schaffalitzky and A. Zisserman. Viewpoint invariant texture matching and wide baseline stereo. In *Proc. 8th International Conference on Computer Vision, Vancouver, Canada*, July 2001.

[8] F. Schaffalitzky and A. Zisserman. Viewpoint invariant texture matching and wide baseline stereo. In *Eighth Int. Conference on Computer Vision (Vancouver, Canada)*, 2001.

[9] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *PAMI*, 19(5):530–535, May 1997.

[10] Cordelia Schmid and Roger Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–535, 1997.

[11] C.-H. Teh and R. T. Chin. On image analysis by the methods of moments. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 10(4):496–513, 1988.

[12] Dennis Tell and Stefan Carlsson. Wide baseline point matching using affine invariants computed from intensity profiles. In *ECCV (1)*, pages 814–828, 2000.

[13] P.H.S. Torr and A. Zisserman. Robust parameterization and computation of the trifocal tensor. In *BMVC96*, page Motion and Active Vision, 1996.

[14] T. Tuytelaars and L. Van Gool. Content-based image retrieval based on local affinely invariant regions. In *Proc Third Int'l Conf. on Visual Information Systems*, pages 493–500, 1999.

[15] T. Tuytelaars and L. Van Gool. Wide baseline stereo based on local, affinely invariant regions. In M. Mirmehdi and B. Thomas, editors, *Proc British Machine Vision Conference BMVC2000*, pages 412–422, London, UK, 2000.

[16] Tinne Tuytelaars and Luc J. Van Gool. Content-based image retrieval based on local affinely invariant regions. In *Visual Information and Information Systems*, pages 493–500, 1999.

[17] J. Wood. Invariant patter recognition: A review. *Journal of the Pattern Recognition*, 29(1):1–17, 1996.

[18] Z. Zhang, R. Deriche, O. D. Faugeras, and Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78(1-2):87–119, 1995.